

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 01:21:45 : Search time 3118 Seconds
(without alignments)
5684.288 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPHAY.....IAYCHSDKIRWHISLGRF 609

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cgn2_1/USFTO.spool/US09857669/runat_05112002_110007_7724/app_query.fasta_1.775
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human4.0.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669.ecgn_1.1879 @runat_05112002_110007_7724 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_wa:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgt_others:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	3173	100.0	1830	6	AX026688	Sequence	AX026688 Sequence
2	3173	100.0	1830	6	AX026690	Sequence	AX026690 Sequence
C 3	3153	99.4	340806	1	NMA122491		AL162752 Neisseria
4	3136	98.8	12429	1	AE002561		AE002561 Neisseria
5	3136	98.8	172325	6	AX044035		AX044035 Sequence
6	3019	95.1	1764	6	AX026692		AX026692 Sequence
C 7	596	18.8	189050	1	AL646069		AL646069 Ralstonia
C 8	449	14.2	11158	1	AE004682		AE004682 Pseudomon
9	384.5	12.1	1374	6	AX078576		AX078576 Sequence
10	380.5	12.0	243184	1	AP002568		AP002568 Escherich
11	379.5	12.0	10819	1	AE000493		AE000493 Escherich
12	379.5	12.0	11760	1	AE000493		AE000493 Escherich
13	379.5	12.0	338534	1	ECOUW93		U14003 Escherichia
C 14	374	11.8	1737	6	AX033468		AX033468 Sequence
C 15	374	11.8	11833	1	U32752		U32752 Haemophilus
16	366.5	11.6	10831	1	AE013668		AE013668 Versinia
C 17	366.5	11.6	216050	1	AJ414157		AJ414157 Versinia
18	357	11.3	24892	1	AE008906		AE008906 Salmonell
19	357	11.3	249050	1	AL627283		AL627283 Salmonell
20	355.5	11.2	1731	6	AX033470		AX033470 Sequence
C 21	348	11.0	14390	1	AE004323		AE004323 Vibrio ch
22	342.5	10.8	201050	1	AL646064		AL646064 Ralstonia
23	342	10.8	333800	1	SME591792		AL591792 Sinorhizo
C 24	336	10.6	10302	1	AE006218		AE006218 Pasteurel
C 25	323.5	10.2	10029	1	AE008175		AE008175 Agrobacte
26	323.5	10.2	10301	1	AE009208		AE009208 Agrobacte
27	306.5	9.7	10074	1	AE003957		AE003957 Xylella f
C 28	304.5	9.6	10029	1	AE009623		AE009623 Brucella
C 29	300	9.5	39013	1	AF088896		AF088896 Zymomonas
C 30	292	9.2	20310	1	AF010120		AF010120 Xanthomon
C 31	291.5	9.2	10839	1	AE012071		AE012071 Xanthomon
C 32	291	9.2	11870	1	AE012531		AE012531 Xanthomon
C 33	283	8.9	334520	1	AP003588		AP003588 Nostoc sp
34	272.5	8.6	10301	1	AE005835		AE005835 Caulobact
35	260.5	8.2	2379	6	NGU81959		U81959 Neisseria 9
36	256.5	8.1	2379	6	AX155443		AX155443 Sequence
37	256.5	8.1	2379	6	AX155443		AX155443 Sequence
38	253.5	8.0	2394	6	AX155447		AX155447 Sequence
39	253.5	8.0	2394	6	AX202497		AX202497 Sequence
C 40	253.5	8.0	340806	1	NMA122491		AL162752 Neisseria
C 41	253.5	8.0	347750	1	AP002998		AP002998 Mesorhizo
42	253	8.0	2394	6	AX155439		AX155439 Sequence
43	253	8.0	2394	6	AX202489		AX202489 Sequence
C 44	253	8.0	13538	1	AE002375		AE002375 Neisseria
C 45	253	8.0	349980	6	AX043922		AX043922 Sequence

ALIGNMENTS

RESULT 1

AX026688
LOCUS AX026688 1830 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent WO0034480.
ACCESSION AX026688
VERSION AX026688.1 GI:10187860
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
1. (bases 1 to 1830)
REFERENCE
AUTHORS Ruelle,J.B.
TITLE meningitidis and polynucleotides bas040 from neisseria
polynucleotides
JOURNAL Patent: WO 0034480-A 1 15-JUN-2000;
RUELLE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
Source
Location/Qualifiers
1..1830
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 440 a 503 c 461 g 326 t
ORIGIN

Alignment Scores:
Pred. No.: 1.17e-204 Length: 1830
Score: 3173.00 Matches: 609
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-857-669-2 (1-609) x AX026688 (1-1830)

QY 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePhePheProHisAlaTyr 20
Db 1 ATGATGATCAAAACCGCCGCTCTCTCGCGGCTTATTTTCTTCGCGACGCATAC 60

QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 61 GCGCCTGCGCGGACCTTTCGGAACAAAGCGCGGGTTTCGCAATGTTCAAAAACAA 120

QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 121 AGCCCGACACCGAATCAGTAAATTAATTAACCCAAATTCCTCGCGCATCCACCGCAG 180

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 181 GATAGTGAATCAAGATATGTCGAAGAACACCTGCGGTATCATCAGCAGCAGGAA 240

QY 81 GluValIleAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 241 GAAGTATTGGACAGAAACACCGGGCTTCTCGCGAAGACCGCGCAACCGTTAAA 300

QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 301 ACCATGCTCCGCGCAAGAGCTATTTCAGCAGCAAAAGTCAGCTGACGGAACACGGA 360

QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 361 GCTTATACGTTACATCATCACCGGGCGCGCCACCAAAATCGCCAACTCGGTGTCGCC 420

QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 421 ATCTCTCGCGCATCTCTTTCAGAGCGCAAGCTCGCGGAATCTACCGCAACGCGTGA 480

QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGluAspSerTrpGluAsnSerLysThr 180
Db 481 AACTGGCAGCAGCCGCTAGGCGATGATTTTCATCAGACAGCATGTTGGGAACACGAA 540

QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 541 TCCGCTCTCGCGCGGTAAACGGAAGAGGTACCGCTTGCCAAAGCTCGCAACACCGCG 600

QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 601 GCGGCGGTCAACCCCGATACCGCCACCGCTGATTTGAACGTCTGTCGACAGCGCGCG 660

QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 661 CCCATCGCTTCGCGGACTTTTGAATCACCGGCACACAGCTTACCCGCAACAAATCGTC 720

QY 241 SerGlyLeuAlaAspPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 721 TCCGCGCTGGGCGGCTTCCAAACCGGCACCGCTTACGACTCGACTGCTGCTGCTGCTTC 780

QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 781 CAACAGGCACTCGAACAAACAGCGCATTTATTCGCGCGCTCGGTACAAACCCGACTTCGAC 840

QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 841 CCGTCTCCAAAGCGACCGCGTCCCGTCAAAAGTCAGCGTAACCGAGGTCAACAGCCACCAAG 900

QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
Db 901 CTCGAACCGGCATCCGCTCGATTTCGGAATACGGTTTGGGCGCAAAATCGCTACGAC 960

QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 961 TATTACAACCTCTTCAACAAAGGCTATATCGGCTCGGCTCGCTCGGATATGACAAATAC 1020

QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 1021 GAAACACGCTTGGCGCGGCATCAGCAGCGCACTATTCGGGGCAACTACTTGGACA 1080

QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 1081 AGCAACGTTTCTACACCGTTTCAGCCACCAACCTCGAAACGCGCTTCTCCGCGC 1140

QY 381 GlyIleTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 1141 GGCATCTGTTATGTCGCGACCGCGCGGCATCGATCCAGGCTGGGGCGCAGATTTCTC 1200

QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 1201 GCAGAGCCCGGAAATCCCGGCTCGGATATCGATTGGGCAACAGCCACCCACGATG 1260

QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
Db 1261 CTCACGCTCTTGGAAACGCGAGCTGCTCAACACGCTGCTGATCCCGAAACCGCAT 1320

QY 441 TyrLeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 1321 TACCTTCGACGGAATATCGGTACGACTTTGGCACATCTCTGCTCCACCGCTGATC 1380

QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLeuGlyThrPhe 480
Db 1381 CGCAGCTCTGCGCGGTGCAAGGTATTCTTCAGCGCGGAAACAAACAACTCGCGACGTC 1440

QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 1441 ATCATACCGGACAAAGCGGTTACCGCTTCCCGCGACANTGCCGACCTCTCTTCAGGG 1500

QY 501 LeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 1501 CTCGATCTCCGACGCGCGGCGCTCTTCGCTGCGGGTTCAGAACTCGACAGCATCGGA 1560

QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGlu 540
Db 1561 CTTGGCGCGCGAAGGATCGGTCTGCCGAACGCGCCCTCTCTGTTGGGAGCCTGGAA 1620

QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValIlePheHisAspMetGlyAspAla 560
Db 1621 TACCAACTGCGCTTACGCGCACCTTTTCGGGCGCGGTGTTCCACGATATGGCGATGCC 1680

361	GCATTATACGGTACACATCACACCGGGCCGGCACCAAAATGCCAAAGTCGGTGTGC	420	Db
141	ITleuGlyAspIleLeuSerAspClyAsnLeuAlaGlyTyrTyrArgAsnAlaLeuGlu	160	Qy
421	ATCTCGGGCAGACATCTTTTCAGAGCGCAACTCGCCGAATACTACCGCAACGCGCTG	480	Db
161	AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr	180	Qy
481	AACGGGAGCAGCGCGTAGGCGAGTGATTTTCATCAGACAGTTGGGAACAACGCAAACT	540	Db
181	SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg	200	Qy
541	TCCGTCTCGCGCGGTACGCGCAAAAGGCTACCGCGTTGCCAAGCTCGCAACACCGCG	600	Db
201	AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg	220	Qy
601	CGGCGCTCAACCCCGATACCGCACCTCGATTGAAGCTGCTGTGACAGCGGGCCG	660	Db
221	ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal	240	Qy
661	CCCATCCGCTTCGGCGACTTTCGAATACACCGGCACACAGCGTTACCCGCAACAAATGCT	720	Db
241	SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPhe	260	Qy
721	TCCGGCTTCGGCGCTTCACACCGGGCACGGCCCTAGACCTCGACTGCTGCTGCATCTC	780	Db
261	GlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp	280	Qy
781	CAACAGGCACTCGAACAAACGGGCATTATTTCGGCGCGCTCGCTCAAGCGCGAGTTG	840	Db
281	ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys	300	Qy
841	CGTCTCAAGGCGACCGCGTCCCGCTCAAGTCAGCGTAACCGAGGTCAACGCCACAAG	900	Db
301	LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp	320	Qy
901	CTCGAAACCGGGCATCCGCTCGATTTCGGAATACGGTTTGGGGGCAAAATCGCCTAG	960	Db
321	TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr	340	Qy
961	TATTACAACCTCTTCAACAAAGGCTATATCGGCTCGTCTGGGATATGCAAAATAC	1020	Db
341	GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr	360	Qy
1021	GAACACCAAGCTTCCCGCGGCATCAGCAGCGCGCAACTATCGGGGCAACTACTGGA	1080	Db
361	SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly	380	Qy
1081	AGCAACGTTTCTTACAACCGTTCGACCACCCCAAACTCTGAAAACGCGCCTTCTCC	1140	Db
381	GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu	400	Qy
1141	GGCATCTGGTATGTGGCGACCGCGGGGCATCATGCCAGGCTGGGGGCAAGTTTCTC	1200	Db
401	AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet	420	Qy
1201	GCAGAGGCGCGAAAATCCCGCGCTCGCATATCATGTTTGGGCAACAGCCACGCCAC	1260	Db
421	LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis	440	Qy
1261	CTGACCGCCTCTTGGAAAGCCAGCTGCTCAACAACTGCTGATCTCCGAAACGGCGAT	1320	Db
441	TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle	460	Qy
1321	TACCTCGACGCGCAAAATCGGTACGACTTTGGGCACATTCCTGTCTCCACCGCGT	1380	Db
461	ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe	480	Qy
1381	CGCACCTTCGCCGCTGCAGGTTATTTCCTTCAGCGCCCGAAACAAACAACTCGGCA	1440	Db
481	IleIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly	500	Qy
1441	ATCATACGCGGACAAGCGGTTTACACGTTCCCGCGCACTTCCGCGCACTTCCGTCAGG	1500	Db

QY 501 LeuMetPheArgSerGlyAlaSerValArgGlyTyrGluLeuAspSerilegly 520
 Db 1501 CTGATGTTCCCGACGGCGCGCTTCTCGTGGCGGTACGAACCTCCACAGCATCGGA 1560
 QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
 Db 1561 CTTGCGCGCCGAAACGATCGTCTCGCCGAAACGCGCTCTCTGCTGGCAGCGCTGGA 1620
 QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
 Db 1621 TACCACATGCGCTTACCGGACCGCTTTCGGCGCGGTTCACAGATATGGCGCATGCC 1680
 QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
 Db 1681 GCGCGCAATTTCAACAGCTATGAAGCTGAACACGCTTCGGGACTGGCGCTGGCTGTTTC 1740
 QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
 Db 1741 AGCCCGCTTGGCGCGTCTTCTTCGACATCGCTACGGCAGCAGGATAAGAAAAATCGC 1800
 QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
 Db 1801 TGGCAGATCAGCTTGGGAACAGCGCTC 1827

RESULT 3
 NMA122491/c 340806 bp DNA linear BCT 04-DEC-2000
 LOCUS NMA122491
 DEFINITION Neisseria meningitidis serogroup A strain 22491 complete genome;
 segment 1/7.
 ACCESSION AL162752 AL157959
 VERSION AL162752.2 GI:7378778
 KEYWORDS
 SOURCE Neisseria meningitidis 22491.
 ORGANISM Neisseria meningitidis 22491.
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.

REFERENCE 1 (bases 1 to 340806)
 AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C., Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T., Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N., Holtroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K., Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M., Skelton,J., Whitehead,S., Spratt,B.G. and Barrick,B.G.
 TITLE Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 340806)
 AUTHORS Parkhill,J.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

COMMENT
 Notes:
 Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES
 source
 1..340806
 /organism="Neisseria meningitidis 22491"
 /strain="22491"
 /db_xref="taxon:122587"
 /note="serogroup: A"
 complement(38..358)
 /gene="NMA0001"
 complement(38..358)
 /gene="NMA0001"
 /note="NMA0001"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein NMA0001"

gene
 38..358
 /note="NMA0001, unknown, len: 106 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein NMA0001"

CDS
 38..358
 /note="NMA0001, unknown, len: 106 aa; similar to hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa), fasta scores: E(0; 80.5% identity in 190 aa overlap, and to proteins involved in cell cycle e.g. TR:O54679 (EMBL:AF036487) Lactococcus lactis plasmid pNZ4000

/protein_id="CAB83321.1"
 /db_xref="GI:7378779"
 /db_xref="SPTREMBL:O9JX93"
 /translation="MSNTQTIHSADFTTTSRAWGALDIANNMGTTVRLUHTDQYKWHV
 NDGEFVAVNDGEVDHMYRNGEHEIVRLKSGDIFYAGICTEHWAPRGTIRLVIEK
 EGSV"
 complement(414..2438)
 /gene="nuoL"
 complement(414..2438)
 /gene="nuoL"
 /EC_number=1.6.5.3
 /note="NMA0002, nuoL, NADH dehydrogenase I chain L, len: 674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored_q1, NADH-ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored_q1_N, NADH-ubiquinone oxidoreductase (complex I), chain 5 N-terminus"
 /codon_start=1
 /transl_table=11
 /product="NADH dehydrogenase I chain L"
 /protein_id="CAB83322.1"
 /db_xref="GI:7378780"
 /db_xref="SWISS-PROT:O9JX92"
 /translation="NNDMTLXIIALVPLAGSLIAGLPGNKIGRAGATTVTILGVAVS
 AVLSAYVLGELNGSPAKFDENVYTLTMGGGLDSVGLVDYTMAMWVTVGVSLMV
 YFKRPSATFANUKAFILNRVGDGFLGLGLVLAIFGSGSLRYQDVFAFLPNVQNTIQ
 LFPVMSITVTCLLLFVGAKGSAQFPLHWLPDSMEGPTISALIIHAATWVTAGL
 FMYRSMSTVSTALSYIVIGAITALFMGLFGVQNDIKRVVAYSTLSQLGYATV
 ALGASAYSVAMFVMTWHPKALFLAAGSAIIGMHIDQDMHMGNLKMYPIWLTMT
 LIGLSLIGTPFSGFYSKDSIIKAYSTLPGSGYAFVAVLSVTFYAFVQVPM
 VFHEKRSKLEPHSDHGHGHEHGLKNDNPESPLVTLPLILLVAPSVIIGYIAI
 EPMYLGDFPKDVIYNADAHPTMHMKKEEFUGALAMVSHSLHSPVLVLAAGLSNWL
 LYVLPFLPAKTAQAFRPYVLFENKYYLDALIFNFAKGTALGCTFFWKGDTALID
 NGIVSGSLVGAVAQAQVRKVTGTFITYTAAWFGVLVLGLMTFWGLFR"
 1035..1044
 /note="Core DNA uptake sequence: gccgtctgaa"
 /label=DUS
 complement(1116..2030)
 /gene="nuoL"
 /note="Pfam match to entry PF00361 oxidored_q1
 NADH-ubiquinone/plastoquinone (complex I), various chains,
 score 351.80, E-value 7.3e-102"
 complement(2061..2243)
 /gene="nuoL"
 /note="Pfam match to entry PF00662 oxidored_q1_N,
 NADH-ubiquinone oxidoreductase (complex I), chain 5
 N-terminus, score 77.20, E-value 3.5e-19"
 complement(2487..2828)
 /gene="NMA0003"
 complement(2487..2828)
 /gene="NMA0003"
 /note="NMA0003, unknown, len: 113 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein NMA0003"
 /protein_id="CAB83323.1"
 /db_xref="GI:7378781"
 /db_xref="SPTREMBL:O9JX91"
 /translation="NMDKQLSQEERKAMLNITYQALNLPQYKATRFQIVNEFGK
 EAADKLLSTGEKKTTQTGTTELISGGVHALKYSMEYLVLPQKMWCDLFTBEQLAVAK
 RLKRVGFVSPK"
 complement(2833..2837)
 complement(2952..3527)
 /gene="NMA0004"
 complement(2952..3527)
 /gene="NMA0004"
 /note="NMA0004, unknown, len: 191 aa; similar to
 hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa),
 fasta scores: E(0; 80.5% identity in 190 aa overlap,
 and to proteins involved in cell cycle e.g. TR:O54679
 (EMBL:AF036487) Lactococcus lactis plasmid pNZ4000

putative mobilization protein (200 aa), fasta scores; E():
0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell
filamentation protein Fic (200 aa), fasta scores; E():
0.093, 29.4% identity in 126 aa overlap
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0004"
/protein_id="CAB83324.1"
/db_xref="GI:7378782"
/db_xref="SPTREMBL:O9JQ99"

/translation="MPSENPIGKTKMSIDEOSLNNARRLPESGDIDRIEIVGTAGLQQ
IHRYPGLYDAGQIDNDSIKGKGFANAMYLKEALVRIEQMPERTEEIIAKYVE
MNIHAPFEGNSRSTIWLQDLVKNLKKVYNQWNSKTYLQAMPSPVNDLEURL
LKDNITDDVDNREIIFKGIQSYYYEGYK"
complement(3516..3525)
/gene="NMA0004"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS..

misc_feature

gene
complement(3561..3866)
/gene="nuoK"

gene

CDS
complement(3561..3866)
/gene="nuoK"

/EC_number="1.6.5.3"

/note="NMA0005, nuoK, NADH dehydrogenase I chain K, len:
101 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
1.1e-24; 65.1% identity in 100 aa overlap. Contains pfam
match to entry PF00420 oxidored_q2.

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
/codon_start=1

/transl_table=11

/product="NADH dehydrogenase I chain K"

/protein_id="CAB83325.1"

/db_xref="GI:7378783"

/db_xref="SPTREMBL:O9JQ99"

/translation="MITLTHLVGLGALLFGISAMGIFNNKKNVLVLMLSTELMLLVN
FNFIAPSHLGTADGQIFVFFVTVAESAIGLAIMLVYRNROTINVADJDELK"
complement(3564..3863)
/gene="nuoK"

misc_feature

/note="Pfam match to entry PF00420 oxidored_q2.

NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
score 141.50, E-value 1.5e-38"

complement(3863..4534)

gene

CDS
complement(3863..4534)
/gene="nuoJ"

/EC_number="1.6.5.3"

/note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fasta scores; E():
9.7e-21; 41.3% identity in 201 aa overlap. Contains pfam
match to entry PF00499 oxidored_q3.

NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1

/transl_table=11

/product="NADH dehydrogenase I chain J"

/protein_id="CAB83326.1"

/db_xref="GI:7378784"

/db_xref="SPTREMBL:O9JX90"

/translation="MTFQILIFIAVILYGAIKTVYAKNPVHAALHLVLFTFCVSM
LWLMQAEFGVTVLVVYVGVAVMLFVVMNLNIDIEEMRAGFRWRHAPVAGVYGTLL
AVALLIYLNPTKIDLAAGKMDIPADYNNTDLGSRITDYDLLPFELAAVLLLLGNV
AATALYTRVTPKRPMDPQVKVRADQGRMLRVMEAVKPTQESAEESEVSDLLKPK
EECKA"

complement(3874..3879)

RBS

/gene="nuoJ"

complement(4034..4525)

misc_feature

/note="Pfam match to entry PF00499 oxidored_q3,

NADH-ubiquinone/plastoquinone oxidoreductase chain 6,

score 77.90, E-value 2e-19"

gene

complement(4570..5277)
/gene="NMA0007"

CDS

complement(4570..5277)
/gene="NMA0007"
/note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"

Alignment Scores:

Pred. No.: 2,48e-200 Length: 340806
Score: 3153.00 Matches: 605
Percent Similarity: 99.51% Conservative: 1
Best Local Similarity: 99.34% Mismatches: 3
Query Match: 99.37% Indels: 0
DB: 1 Gaps: 0

US-09-857-669-2 (1-609) x NMA1Z2491 (1-340806)

QY 1 MetMetLleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
DB 289737 ATGATGATTAAACCGACCGCCTGCTCCTCGCGCTTATTTTCTTCCGACGCATAC 289678
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
DB 289677 GCGCTGCCGCGGACCTTTCGGAACACAGCGCGGGTTCGCATTGTTCAAAACAAA 289618
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgLileAspThrGln 60
DB 289617 AGCCCGACACCGAATCAGTTAAATTAACCCCAATTCCTCGCTCGCATCGACACGAG 289558
QY 61 AspSerGluLleLysAspMetValGluGluHisLeuProLlePheThrGlnGlnGlnGlu 80
DB 289557 GATAGTGAATCAAGATATGTCGAAGAACACCTGCCCTCATTCACGACGACGAGGAA 289498
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
DB 289497 GAAGTATTGGACAAGAACACGCGGCTTCTCGCGAAGAACACCGCACACGTTAAA 289438
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
DB 289437 ACAATGCTCCGACCAAGGCTATTTACGACGCAAAAGTCAGCTGACGGAAAAACACGA 289378
QY 121 AlaTyrThrValHisLleThrProGlyProArgThrLysLileAlaAsnValGlyValAla 140
DB 289377 GCTTATACGTTACACATCACACCGCGCGCCGCAAAATCGCAACGTCGCGCTCGCC 289318
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyr-TyrArgAsnAlaLeuGlu 160
DB 289317 ATCCTCGCGGACATCCCTTTCAGAGGCAACCTCGCGGATACTACCGCAACCGCTGGAA 289258
QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
DB 289257 AACTGGCAGCAGCGGTAGGCAGTATTCGATCAGCAGAGTTGGGAAACAGCAAACT 289198
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
DB 289197 TCCGTCCTCGCGGGGTACGCGCAAGGCTACCCCTTGCACAGCTCGGCAACACCGGG 289138
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
DB 289137 GCGGCGGTCAACCCCGATACCGCCACCGCGGCGGATTTGAACGTCGTCGTGACACGCGCGC 289078
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
DB 289077 CCCATCGCCTTCGGGACCTTTGAAATTTACCGCACGCGGTTACCCCGCAACAAATCGTC 289018
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
DB 289017 TCCGCGTTGGCGCGCTTCCAAACCGCGCACGCCCTTACGACCTCGACCTGCTGTCGACATC 288958
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
|||||


```

/db_xref="GI:7227386"
/translation="MIPYKRNPMKEYKVIYIQSLLSLFFGCAAKVNPIKSEFLNK
QTPGWRVVTMEKDRMLLFKREAYVILEROV"
1813..2028
/gene="NMB2131"
CDS
1813..2028
/gene="NMB2131"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF42439.1"
/db_xref="GI:7227387"
/translation="MFKLGVTCLGLFAGWLLLVOLWFSLEAELEFKKILTWAGL
FVILAAALVCGQYSEKKMKDGFN"
complement(2104..3570)
/gene="NMB2132"
CDS
complement(2104..3570)
/gene="NMB2132"
/notes="similar to PID:747922 percent identity: 48.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transferrin-binding protein-related protein"
/protein_id="AAF42440.1"
/db_xref="GI:7227388"
/translation="MFKRSVTAMACIFALSACGGGGSDPKSADTLSPKPAVPVSE
KTEAEADPAQSGOCGAPSAQGSODMAAVENTNGGAVTADNPKNEDVAQNDM
PQNAQETDPSHPDPNMLNAGNENQATDAGESSQAPQPDMAADCMQGDGDSA
GQNAQNTAAGCAAGNAGAGSDP IPASNAPANGSGNFCRVLDANGVLVADGVSQ
NITFLTKGDSGNGNEDEEVLKSEPEKLSADAKISNYKDKGNDKFLVGLVADSVQ
MKGINQYLIIFYPKPTSPARFRSARSRLPAEMPLIPVNOADTLVIGRAVSLTGH
SGNIPAPENYRXYLTGAELKPGSYALRVGEPKAGEMLAGAAYNGEVHLHTENG
RPTPTGRPAKVDYGSKSDGDIIDSGDLDHMGTKFKAAIDGNGFKGTWTEGNSGDY
SGKFGPAEEVAGYSYRPTDAEKGFGFVAGKKEQD"
complement(3825..5054)
/gene="NMB2133"
CDS
complement(3825..5054)
/gene="NMB2133"
/notes="similar to SP:P42602 PID:606028 GB:U00096
PID:1789473 percent identity: 82.74; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="sodium/dicarboxylate symporter family protein"
/protein_id="AAF42441.1"
/db_xref="GI:7227389"
/translation="WAFGKSLFPHATGRVSLVROIAAGLALGIVIGSVSPOLGLAAGLF
GSLFVCAKAVAPLVLTVAATIAQHKGNKAHRIPIVLYLIGTFSAALTAVIAGM
VFTHVLVAGAGDVAAPPSGIVELKSLNLANVAINANAYIGILLAWALVGA
ALRNHGSDFTRQVADLAENVSTVVKWIRIAPLIGLGLVSSTIAETFGGALYAKL
LAVLLCGMFIALANPAINVMMKIRNPFPLVFTCLRESGVYAFFTRSSAANTPVNNA
LAKKGLHEDTYSISIPLGATINMAGAAITITVLMAAAHTQGITVDFAILLLSVA
TVSACGASGAGVSGSLILIPLACSLFGIDNVAMQVAVGFIIGVIOQDSAEALNSFTD
VLFTAAADLGRORNAE"
5527..7434
/gene="NMB2134"
CDS
5527..7434
/gene="NMB2134"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42442.1"
/db_xref="GI:7227390"
/translation="MPSEASREVRTVLKSKAATPMHDTRTMMIKPTALLPALFFPPH
AYAPADLSENKAAGFALFKNKSPTDESVKLKPFPVLIDTQDSEIKDMVEEHLPLIT
QOOEVEDKQEGFLAEAPVNTMLRSKGYFSKSVSLTEKDGAVTVHITPGPRKI
ANYGVAILDLSGDNLADYRNALNWOQVPDSDFQDQSWENSKTSVLGAVTRKAP
LAKLGNTOAANPDATADLNVVDSGRPIAGFDLGTGTQVRPEQIVSGLARFQPM
PYDLDDLDFQQALEQNGHSGASVQADFDRLDGRVPRVSVTEVRHKLKLEIGRLD

```

```

SEYGLCKLAIYDYYNLFNKGYIGSVVDMDKYETTLAAGISOPRNYKGNWTNSVSN
RSTQLEKRAFSGGVYVYDRAGIDARGAELFAGREKPIPGSADLGNHATMLTAS
WKQLLNNVLPENGHYLDGKTGTTGLTGLTSTLIRT SARAGYFFTFPHKILGTI I
RGOAGYTVARDNAVPSGLMFRSGASSVGRYELDSICLAGPNSGVYPERALLVGSLE
YQLPPTRLSGAVFHDMDGDAANFKRMLKHGSGLVGRWFSPLAPPSFDIAYGHSKK
IRWHISLGRF"
7441..11658
/gene="NMB2135"
CDS
7441..11658
/gene="NMB2135"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF42443.1"
/db_xref="GI:7227391"
/translation="WATSDIAAHFETDITWTDAPTDTDTENGTRKMSHRPTPP
AKRRPLKLKLSAALLSVLIVCFGLWAGTAGLRFGLYOIPSWFGVNSSONLKG
LDPFGDGNMSTETEGADLKISRFPAWKPSMLRRSLHTEISAGDIAIVTKPTPK
EERPPLSDPSIDLPAAYVLDRFETGKISMGKAFDKQTVYLERLDASRYDRKGRHLD
LKAADTPWSSSGAASVGLKDPFALDTAITYKGLGKTIHSTARLSGLKDVRAELA
IDGNTIRLSGKSVIHPFAESLDKTEEVKGFENINPAAFVPSLPDLAGLPDLTAIPS
FSDGIALGSLDLINTKAGFADRNCPVROVLGGFVIRODGTVHIGNTSAALGRGCI
RLSGKIDTEKDLIDLINIGINSVGAEDVLQTAFKGRLOGSIGGTTSPKLSHQLGIG
TARTDSLAISADPANGQRKLVDITVNAAGQSGLSAQGLLELFRDLKLDLNSRAF
DPSRIDPQLPAGNINGSINLAGELAKEPTGMRPLPGTFNGVPIAGSADIVYESRHL
PRAVRLRGRNIITDGGFGKGRKLNLTITPDLRSFGFGLAGSLNVRHLSGDDLD
GVTTFETDLSGAARNLHIGKAADIRSLDTFKGSPDTSRIRADIRGSRLSGGAA
VWOTADMLDGTGVHRIRTHAAMTLDGKPFKFDLDASGGINRELTRWKSIGITDIG
GAFNLKONMRTLEAGARVAASAAWMOAGSLNLOHFSWDKGTGISAKGGAHGLHI
AELHNFKPPFPHNLVLDVAVGRNARCYLANISQSDGDLVLPGGOALGNFSLK
TRQNDRIGILLDGGARFRINADGLIANAGGGMANAPLGGRTIASLPDLGALKPPL
PAAQNTIGSLNAAQIGRGVGSVNAAVTGGSIADPHLGSGINDSKLYIRNQTGG
RLMTVADAEVFNFLPVGQTVKGSNAAVTGGSIADPHLGSGINDSKLYIRNQTGG
IILDNGLSHIAGRWKVIIDSLKFRHEGTAEISGTVGNENSPDVIDGAVDFYRILS
RPNRLTVSGNTRLYSPQKISVTGMITDQGLFGQKSSMPSGVDGVDVYLVAGEVKE
AAAPLPVNMNLTLDLDGIRFAGYADVTIGKLTLTAAQSGSVRGVTVRVIKRGYK
AYGDOLDITKGTVSFVPLNDPLNIRAERLSPVAGVEITLGSLSPPRITLTANPEM
SEKDKLSWLTINRAGSGSDNAAASAAGALLAGQINDRIGLVDLGGFTSKSRNAQ
TGLNLAQEVLTWVGQLTGKLYGYEYSISSAESQSVKLYLTRAITRAQVARISSRG
GEUTYTRFRFSGSKDKDSAGNGK"
BASE COUNT 2986 a 3768 c 3144 g 2531 t
ORIGIN
Alignment Scores:
Pred. No.: 4,48e-201 Length: 12429
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 1 Gaps: 0
US-09-857-669-2 (1-609) x AE002561 (1-12429)
Qy 1 MetMetileLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTry 20
|||||
Db 5605 ATGATGATCAACACCGACCGCCCTCTCTCCCGGCTTATTTTCTTCGGCAGCATAC 5664
Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
|||||
Db 5665 GCGCCTGCGCGCGACCTTTCGAAAAACAAGCGCGCGGTTTCGCATTGTTCAAAAACAA 5724
Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgileAspThrGln 60
|||||
Db 5725 AGCCCCGACACGGAATCAGTCAAAATTAACCAATATCCCGCTCTCATCGACGAG 5784
Qy 61 AspSerGluileLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGln 80
|||||
Db 5785 GACAGTGAATCAAGATATGGTCAAGAACACCTGCCGCTCATCAGCAGCAGCAA 5844
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
|||||

```

Db 5845 GAAGTATTGGACAAGAACACAGCGGCTTCTCGCCCAAGAACGCGCGGACAACTTAAA 5904
Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 5905 ACAGTGTCTCCGCGAGCAAGAGGTATTTTCAGGACGAAAGTACGCTGACGGAAAGACGGA 5964
Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 5965 GCTTATACGGTACACATCACACCGGCGCGCCGACCAAAATCGCCAACTCGCGCTCGCC 6024
Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 6025 ATTCCTCGCGACATCTTTTACAGACGGCAAGCTCGCGGAATACTACCGCAACCGCTGGA 6084
Qy 161 AsnTyrGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 6085 AACTGGCAGCAGCGGTAGGACGAGTTCGATCAGACAGTTCGGGAAACAGCAAACT 6144
Qy 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 6145 TCCGTCTCGCGCGGTAAACGCAAGACCTACCGGCTTCCCAAGCTCGCAATACCGAG 6204
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 6205 GCGGCGGTCAACCCGATACCGCCACCGCGGATTGNACGTCTGCTGACAGCGCGCG 6264
Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 6265 CCATCGCTCTTGGCGAGCTTTTGAATCAGCGGACACAGCGTTTACCCGAAACAAATCGTC 6324
Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 6325 TCCGCGCTTCCGCGGTTCCAGCGCGGTATGCGCTACGACCTCGACCTGCTCGACATTC 6384
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 6385 CAACAGCGCGCTCGAACAAGCGGCAATATTCGCGCGCGTCCGTACAGCCGACATTCGAC 6444
Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 6445 CGCCTCCAAAGGCGACCGCTCCCGCTCAAAAGTCAGGTAAACGAGGTCAACGCCACAAA 6504
Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyClyLysIleAlaTyrAsp 320
Db 6505 CTCGAAACCGGATCCGCTCGATTTCGGAATACGCTTGGCGGCAAAATCGCTACGAC 6564
Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 6565 TATTACAACCTCTTCAACAAAGGCTATATCGGTTTCGCTCGTGGGATATGCAAAATAC 6624
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 6625 GAACACCGCTTGGCGCGGATCAGCAGCGCGCACTATCGGGGCAACTACTGGACA 6684
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 6685 AGCAACGTTTCTTACAAACGTTTCGACCAACCAAACTCGAATAACCGCGCTTCTCCGCG 6744
Qy 381 GlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 6745 GCGCTCTGTATGTGCGGACCGCGGCGGATCGATGCCAGGCTGGGCGGGAATTTCTC 6804
Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 6805 GCAGAAGCGCGGAAATCCCGCGCTCGCTGTCGATTGGGCAACAGCCACGCCACGATG 6864
Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
Db 6865 CTCACCGCTCTTGGAAACGCCGCTCTCAACAACGTCGTGATCCCCGAAACGGCAT 6924
Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 6925 TACCTCGAGCGGCAAAATCGTACGACTTTGGGACATTCCTGTCTCCACCGCGCTGATC 6984

Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 6985 CGCACCTCTGCCCGTGCAGGTATTCTTACGCGCCGAAACAAAACCTCGCAGCTTC 7044
Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaAlaArgAspAsnAlaAspValProSerGly 500
Db 7045 ATCATACGCGGACACGCGGTTACACGCTTGCCTGGCGACAAATGCCGACGTTCTTTCAGG 7104
Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 7105 CTGATGTTCCGAGCGGGCGGCTTCTTCCGTGCGCGGTTACGAATCTCAGACGATCGGA 7164
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 7165 CTTGCCGCGCGAAGCGATCGGCTCTGCCGAAAGCGCCCTCTCTGGTGGCAGCTGGA 7224
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 7225 TACCAACTGCGCTTTTACGCGACCGCTTTCGCGCGCGGTTCACACGATATGGCGATGCC 7284
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 7285 GCGGCAATTTCAACAGTATCAAGCTGAAACACGTTTCGGGACTGGCGCTGGCTGGTTC 7344
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 7345 AGCCGCGCTTGCCTGCTTTCGATCGCTTACGGGACAGCGATGAAGAAATCCGC 7404
Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 7405 TGGCACATCAGCTTGGGAACGCGCTTC 7431
RESULT 5
AX044035 Sequence 114 from Patent WO0066791.
LOCUS AX044035 173235 bp DNA linear PAT 24-NOV-2000
DEFINITION AX044035
VERSION AX044035.1 GI:11342919
KEYWORDS Neisseria meningitidis.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
REFERENCE 1 (bases 1 to 173235)
AUTHORS Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J.C., Masiagnani, V., Galeotti, C., Mora, M., Ratti, G., Scariselli, M., Scarlato, V., Rappuoli, R., Frazer, C.M. and Grandi, G.
TITLE Neisseria genomic sequences and methods of their use
JOURNAL Patent: WO 0066791-A 114 09-NOV-2000;
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
1..173235
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
/note="sequence too long, cut in 8 pieces.-seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 2272325 173235 bases"
BASE COUNT 43072 a 47583 c 41465 g 40205 t
ORIGIN
Alignment Scores:
Pred. No.: 1,41e-199 Length: 172325
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 6 Gaps: 0

US-09-857-669-2 (1-609) x AX044035 (1-172325)

QY 1 MethMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 141935 ATGATGATCAAAACCGACCGCCCTGCTCCCTGCGCGCTTTATTTCCTTCGCGACGCATAC 141994

QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 141995 GCGCCTGCCCGCACTTCCGAAACAAAGGCGCGGCTTCGCGATTGTTCAAAACAAA 142054

QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 142055 AGCCCCGACACCGAATCAGTCAAAATTAACCCAAATTCGCCGCTCATCGACACGAC 142114

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGlu 80
Db 142115 GACAGTCAAAATCAAGATATGTCGAAGAACACCTGCCGCTCATCGCAGCAGCAGGAA 142174

QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 142175 GAAGTATTGCAAGGAACAGACGGCTTCCTCGCGAAGAAGCGCGGACCAAGTTAAA 142234

QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 142235 ACGATGCTCGCAGCAAAAGGCTATTTCAGCAGCAAAAGTCAGCTGACGGAAAAAGACGGA 142294

QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 142295 GCTTATACGGTACACATCACCGGGCGCGCACCAAAATCGGCAACGCTCGCGCTCGCC 142354

QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 142355 ATCCTCGCGGACATCCTTTCAGACGGCAACCTCGCGGAATACTACCGCAACGCGCTGAA 142414

QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 142415 AACTGGCAGCAGCGGTAGCGAGCATTTTCGATCAGGACAGTTGGGAAACAGCAAAACT 142474

QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 142475 TCCGTCCTCGCGCGGTAAACCGCAAAAGCCTACCCGCTTGGCAAGCTCGGCAATACGAC 142534

QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 142535 GCGCGCTCAACCCCGATACCGCCACCGCGATTTGAACGTCGTCGTGGACGCGCGCGC 142594

QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 142595 CCCATCGCTTCGGCGCACTTTGAATCACCGGCACACAGCGTTACCCCGAAACAAATCGTC 142654

QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 142655 TCCGGCTTCGGCGTTCCAGCCCGGTATGCGCTACGACCTCGACCTGCTGCTCGACTTC 142714

QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 142715 CAACAGGCGCTCGAAACAAAGGCGATTTTCGCGCGCGTCCGTACAAGCGGCTTCGAC 142774

QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 142775 GCGCTCAAGCGACCGCGTCCCGCGCAANGTCAGCGTAACCGAGGTCAACACGCCACAA 142834

QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
Db 142835 CTCGAAACCGGCATCGCCCTCGATTCGGAATACGCTTTGGCGGGCAAAATCGCTACGAC 142894

QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 142895 TATTACAACTCTTCAACAAAGGCTATATCGGTCGCTCGGTATCGCAAAATAC 142954

QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 142955 GAAACACGCTTGGCGCGCATCAGCCAGCGCGCAACTATCGGGGCAACTACTGCGACA 143014

QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 143015 AGCAACGTTTCTTACACCGTTTCGACCAACCAAACTCGAAAAACGCGCTTCTCCGGC 143074

QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 143075 GGCCTGTGTTATGTGCGCGACCGCGCGCATCGATGCGAGGCTGGGCGCGAATTTCTC 143134

QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 143135 GCAGAGGCGCGAAATTCGCCGCTCGCTGCGATTTGGGCAACAGCAGCCGACCATG 143194

QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
Db 143195 CTGACCGCTCTTGGAAACGCCAGCTGCTCAACAGCTGCTCATCCCGAAACGCGCAT 143254

QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaIle 460
Db 143255 TACCTCAGCGCAAAATCGGTACGACTTTGGGCACATTCCTGCTCTCCACCGCTGATC 143314

QY 461 ArgThrSerAlaArgAlaGlyTyrPheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 143315 CGCACCTCTGCCGCTGCGAGTTATTCTTTCACGCCCGAAACCAAAACTCGGCAGCTTC 143374

QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 143375 ATCATACGCGCAAGCGGCTTACACCGTTGCCCGCGCAATGCGCGAGCTTCTTCAGGG 143434

QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 143435 CTGATGTTCCGACGCGCGCGCTCTTCGTCGCGGTTACGAACTCGACAGCATCGGA 143494

QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 143495 CTTCCCGCGCGCAACGAGTCCGCTGCGCGCAACGCGCGCTCTGCTGGGCGAGCTCGAA 143554

QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValAlaPheHisAspMetGlyAspAla 560
Db 143555 TACCAACTCGCTTTACGCGCACCTTTCGCGCGCGTTCCTCCACGATATGGCGAGTCC 143614

QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 143615 GCCGCCAATTTCAACGATGAAGCTGAACACGCTTCGGGACTCGGCGTGGCGTGGTTC 143674

QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 143675 AGCCGCTTGGCGGCTTTCTTCGACATCGCTACGGGCACAGCATGAAGAAATCCGC 143734

QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 143735 TGGCAGTACGTTGGGAACGCGCTTC 143761

RESULT 6
LOCUS AX026692 1764 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 5 from Patent WO0034480.
ACCESSION AX026692
VERSION AX026692.1 GI:10187862
KEYWORDS
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
REFERENCE 1 (bases 1 to 1764)
AUTHORS Ruelle, J. L.
TITLE Polypeptides and polynucleotides bas040 from neisseria meningitidis and vaccine comprising said polypeptides and polynucleotides
JOURNAL Patent: WO 0034480-A 5 15-JUN-2000;
RUELLE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES Location/Qualifiers
source 1. 11764

/organism="Neisseria meningitidis"

BASE COUNT 424 a 581 c 451 g 308 t

ORIGIN

Alignment Scores:

Pred. No.: 2,58e-194 Length: 1764
Score: 3019.00 Matches: 579
Percent Similarity: 99.15% Conservative: 3
Best Local Similarity: 98.64% Mismatches: 5
Query Match: 95.15% Indels: 0
DB: 6 Gaps: 0

US-09-857-669-2 (1-609) x AX026692 (1-1764)

QY 23 AlaAlaSerLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerPro 42
|||||
DB 1 GCCGCGGACCTTTCGAAACAAAGCGCGGGTTTCGCATTGTTCAAAACAAAGAGCC 60
QY 43 AspThrGluSerValLysLeuLysProLysPheProValArgLysThrGlnAspSer 62
|||||
DB 61 GACACCGAATCAGTCAAAATTAACACCAATTCGCCGCTCTCATCGACACGAGGACAGT 120
QY 63 GluLeLysAspMetValCucLysLeuProLeuLeuPheLeuLeuGlnGluVal 82
|||||
DB 121 GAATCAAGATATGTCGAAGACACCTGCCGCTCATCCAGCAGCAGGAGGAAGTA 180
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
|||||
DB 181 TTGGACAGGACACAGCGGGTTCTCCGCGAAGAGCGCGGACAACTTAAACGATG 240
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
|||||
DB 241 CTCGCGCAAGAGCTATTTTCAGCAGCAAGCTCAGCGTCCGCGAAGAGAGCGGCTAT 300
QY 123 ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
|||||
DB 301 ACGGTACACATCACACGGCGCGCGGACCAAAATCGCAAGCTCGCGCTCCGCACTC 360
QY 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
|||||
DB 361 GCGCAGATCTTTACAGCGCAACCTCGCGCAATCTACGCAACGCGTGGAAACATGG 420
QY 163 GlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerVal 182
|||||
DB 421 CAGCAGCGGTAGGCGGGATTTTCGATCAGGACAGTTGGGAAACAGCAAACTTCCGTC 480
QY 183 LeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAla 202
|||||
DB 481 CTCGCGCGGTAAACGCGCAAGCGCTACCGCTTCCCAAGCTCGGCAATACGCGAGGCG 540
QY 203 ValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIle 222
|||||
DB 541 GTCAACCCCGATACCGCACCGCGGATTTGAACGTCGTCGTGGACAGCGCGCGCCGATC 600
QY 223 AlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGly 242
|||||
DB 601 GCCTTCGCGGACCTTTGAATCACCAGCACACAGCGTTTACCGCGAACAATCGTCCGCG 660
QY 243 LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGln 262
|||||
DB 661 CTTCGCGGTTTCCAGCCCGGTATCGCCGTACGACCTCGACCTGCTGCTGCTTCCAAAC 720
QY 263 AlaLeuGluGlnAsnGlyHisTyrSerGlyValAserValGlnAlaAspPheAspArgLeu 282
|||||
DB 721 GCGCTCGAACAACAGCGGCAATATTCGCGCGCGTCCGTCACAGCCGACATTCGACCGCTC 780
QY 283 GlnGlyAspArgValProValSerValThrGluValLysArgHisLysLeuGlu 302
|||||
DB 781 CAAAGCGGCGGCTCCCGCTCAAGTCAAGTCAAGGTCAGGTCAGGTCAGGTCAGGTCAG 840
QY 303 ThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyr 322
|||||

DB 841 ACCGCGATCCGCTCGATTTCGGAATACGGTTTGGCGCGCAAAATCCGCTACGACTATTAC 900
QY 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr 342
|||||
DB 901 AACCTTTCAACAAAGGCTATATCGTTCGCTGCTGGGATATGGACAAATACCAAC 960
QY 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
|||||
DB 961 ACGCTTGGCGCGGATCAGCCAGCGCACTATCGGGCACTACTGGACAAAGCAAC 1020
QY 363 ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIle 382
|||||
DB 1021 GTTTCCTACAACCGTTTCGACCAACCAAAACCTCGAAACCGCGCTTCTCCGCGCGCTC 1080
QY 383 TrpTyrValAlaArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu 402
|||||
DB 1081 TGGTATGTGGCGACCGCGCGGATCGATCCGAGCTGGGGCGGAATTTCTCCAGAA 1140
QY 403 GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr 422
|||||
DB 1141 GCGCGAAATTCGCGGCTCGCTGCTGATTTGGCAACAGCCACGATGCTGAC 1200
QY 423 AlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeu 442
|||||
DB 1201 GCCTTTGGAAACGCCAGCTGCTCAACAGCTGCTGCATCCGGAACGGCCATTTACCTC 1260
QY 443 AspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThr 462
|||||
DB 1261 GAGCGCAAAATCGGTACGACTTTGGGCACATTCGTGCTCCACCGCTGATCCGAC 1320
QY 463 SerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIle 482
|||||
DB 1321 TCTGCGCTCGAGGTTATTTCTCACGCCGCAAAACAAACAACTCGGCACGTTTCATCA 1380
QY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
DB 1381 CCGGACAAAGCGGTTACACGTTGCGCGGACAAATCGCGACGTTCTTCAGGGGTGATG 1440
QY 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
DB 1441 TTCCGCGAGCGCGCGCTCTTCGTCGCGGTTACGAACCTCGACAGCATCGGACTG 1500
QY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
DB 1501 GCGCGGAAACGGATCGGCTCGCGAAGCGCGCTCTGCTGGTGGCAGCTGGAAATACCAA 1560
QY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 562
DB 1561 CTGCGCTTTACGCGCACCTTTTCGCGCGGCTGTCACGATATGGCGCATGCCGCGCC 1620
QY 563 AsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
DB 1621 AATTTCAACAGTATGAAGCTGAACAGCGGTTCGGGACTGGCGCTGGCTTCAGCCCG 1680
QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHis 602
DB 1681 CTTGCGCGCTTTCTTCGACATCGCTACGGGCACAGCGATGAAGAAATCCGCTGGCAC 1740
QY 603 IleSerLeuGlyThrArgPhe 609
DB 1741 ATCAGCTTGGGAACACGGTTC 1761

RESULT 7

AL646069/c

LOCUS

DEFINITION

segment 13/19

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL646069 Ralstonia solanacearum GM1000 chromosome, complete sequence;

segment 13/19

AL646069 AL646052

AL646069.1 GI:17429247

Ralstonia solanacearum.

Ralstonia solanacearum

Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

[illegible]

QY 362 AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly 381
Db 5224 GACCTGGTCGACGAGAGCAAGCTGCTGACCCCTT-----GGCGGC 5183
QY 382 IleTyrThrValArgAspArgAlaGlyAlaPheLeuGluGlyAlaGluPheLeuAla 401
Db 5182 GAATGCACAGCAGCGCCGCGATGGTGCAGCGGTGTTCGCTGAACCTGGATCGC 5123
QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu 421
Db 5122 GAGGAGTCAAGCTGGGGCAG-----GACTCGGGGTGAGCAGCTTCCTGTATGCGG 5072
QY 422 ThrAlaSerTrpLysArgGlnLeuAsnValLeuHisProGluAsnGlyHisTyr 441
Db 5071 GGGATCGGCTACTCGCTCTGGAGCCGACCAAGCTCGACCCCGCATGGCTACCGC 5012
QY 442 LeuAspGlyLysIle---GlyThrLeuGlyThrPheLeuSerSerThrAlaLeu--- 459
Db 5011 CTGCAGTTCAAGCTCAAGGGGGGGAAGAGGCTGCTGGCGCAGCCGACGCTCTCCAT 4952
QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
Db 4951 GTCGAGCGCATGGCCAAAGGCTGACAGCTTCGCGGGCGGCATCGCTGCTCGC--- 4895
QY 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaPheValProSer 499
Db 4894 -----CGCTCGAGTAGGCGGAATCGCCACCAACGACTACAGTCTGCTCCGCC 4844
QY 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 4843 TCGCTGGCTCTTCTCGCTGGCGGCGACAGAGCGTSCGTGCTACGACTACCGGACGCTG 4784
QY 520 GlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeu 539
Db 4783 TCGCGGAGAAATTCGATGCGGAGCAAGATCGGGCGGCTTACATGATCCCGCGGACGCTC 4724
QY 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 559
Db 4723 GAGTATCAATATCCGCTGGCGGCGGCTGCGGCTGCGGACCTTCGTCACCGAGGCAAC 4664
QY 560 AlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrp 579
Db 4663 GCGTCAACTCGCTGGACTTCCCTGCTCAAGACCGGGTTCGGCTGCGGCTGCGGCTG 4604
QY 580 PheSerProLeuAlaProPheSerPheAspIleAlaTyrGly---HisSerAspLysLys 598
Db 4603 GTCTCGCGGTCGCGGCTGCGGCTCGACTGCGCCCATGCGCTCGACGACGACGCGGT 4544
QY 599 IleArgTrpHisLeuSerLeuGly 606
Db 4543 TTCCGCTGCACTTCTCCATGGG 4520
RESULT 9
AX078576 1374 bp DNA linear PAT 22-FEB-2001
LOCUS
DEFINITION Sequence 90 from Patent WO0107624.
ACCESSION AX078576
VERSION AX078576.1 GI:13158218
KEYWORDS
SOURCE Pseudomonas putida
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 1374)
AUTHORS Fraser,C.M., Venter,C., Tuemmler,B., Hoheisel,J., Dueterhoeft,A.,
Hilbert,H., Timmis,K.N., Moore,E., Straetz,M., Heim,S.,
Nelson,K.E., Hickey,E. and Peterson,J.
TITLE DNA sequences which are suited for specifically detecting
Pseudomonas putida kt2440
JOURNAL Patent: WO 0107624-A 90 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US); QIAGEN GmbH (DE);
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE);

Deutsches Krebsforschungszentrum (DKFZ) (DE); Medizinische
Hochschule Hannover (DE)
Location/Qualifiers
1. 1374
/organism="Pseudomonas putida"
/db_xref="taxon:303"
BASE COUNT 264 a 428 c 429 g 251 t 2 others
ORIGIN
Alignment Scores:
Prod. No.: 3.41e-17 Length: 1374
Score: 384.50 Matches: 115
Percent Similarity: 42.52% Conservative: 81
Best Local Similarity: 24.95% Mismatches: 224
Query Match: 12.12% Indels: 41
DB: 6 Gaps: 11
US-09-857-669-2 (1-609) x AX078576 (1-1374)
QY 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaVal 186
Db 43 GGTGAACAACATCAATCAGCGGCATTACKAGGATGCAAGCGTTGATCCAGAACCAGCGG 102
QY 187 ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
Db 103 TCGCGTATGGCTTCTTCAGTGGCGCTTCAGTAGCCAGCGCTGCGCGTGCACCGCAA 162
QY 207 ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAsp 226
Db 163 GCGCGTGTGGCGGATATCAAGTGTACAGAGTGGCCGCGTTATCGCTGGCGCG 222
QY 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
Db 223 GTCACCTCGGTGGCGCACACCGCTGGACGAGCACTGCTGCGGCGCATGTGTCGTTC 282
QY 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 283 AAGCGGGGTACCCCTTACGACTCGGAACCTGTCGAGAGCTGAACAACGACCTGCAATCG 342
QY 267 AsnGlyHisTyrSerGlyAlaSerValGluAlaAspPheAspArgLeuGlnGlyAspArg 286
Db 343 ACGGCTATTTTCGAAGCGTGCAGTGGTGGCGGCCCGCCACTGCTGCTGGCGCAAGAA 402
QY 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg 306
Db 403 GTCCCGGTGGATGTTTCATCTGGAACCGTAAACCCAGCCACCATGGCGCTTGGCTGGCG 462
QY 307 LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTrpAsnLeuPheAsn 326
Db 463 TTCCTGACCGAGTCCGGCGCGCGGCGGCAAGCCAACTGGACCCGCGCTGGGTGCAACCCA 522
QY 327 LysGlyTyrIleGlySerValValTrpAspMetAspLys----- 339
Db 523 CAAGGCCAC-----AGCTATGGTGGGAACCAAGCACTGTGCGCGCGCGCCAGAACGTC 576
QY 340 -----TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 577 GGCCTGTGTGTGATACATTCCTCCCTCGACCGCGCTGACCGCAAGTTCGCTTTCGCGCGC 636
QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 637 GGCTACCAAGACGAGGAGCTT-----GCCGCGCACCGACACGCTCAGCAAGCTG 684
QY 377 AlaPheSerGlyGlyIleTyrTrpValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 685 TTGACGGTCCGCGCGCGAGTGGCACAGCAGCTGCCAGTGGCTGGCGCGGGTGATTTCG 744
QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 745 CTCAGTACCAACGAGGAATAATCGCTG---GGTCGACGACTCCGCTTTTGAGCAACCTG 801
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436

Db	802	CTGATCGCGGGGTACGCTTTCTCTCTCGCGCAGT-----GACAACCGGTATCGATCGC	855
Qy	437	GIUASNGlyHisTyrLeu-----AspGlyLysIleGlyThr	448
Db	856	CACAACGGGTATCGCGTGCATGTTCAAGTGCCCAAGGAGCGGTGGTCCGAC	915
Qy	449	ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr	468
Db	916	ACC-----AACCTGTCGACGGCAACGATTGCTCAAGCGCTCGACCACTCGGCCAC	969
Qy	469	PhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla-----	486
Db	970	-----AACACCGCTTTCTCGGCGCTGACAGTTGTTGGTGGCAGTGCACCAAT	1017
Qy	487	GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly	506
Db	1018	GCCTTC-----AAGAACAACATTCGCGCTCGCTTCTTCGCGGT	1062
Qy	507	GlyAlaSerSerValArgGlyTyrCluLeuAspSerIleGlyLeuAlaGlyProAsnGly	526
Db	1063	GGCACCACAGTGTGCGCGTTTACCACTACCAAGACCTGTCGCCGAAGACACGACGT	1122
Qy	527	SerValIleuProGluArgAlaLeuValGlySerLeuGlyTyrGlnLeuProPheThr	546
Db	1123	GACCGTATCGCGGGCGCTACTTGTGTCGACGAGTGTGAGTACCAATTTCGCTGACC	1182
Qy	547	ArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArg	566
Db	1183	GAANAATGCGGGTGCAGCGTGTGTCACCAAGCAACTCGTTCAACGACCTGGAGCTG	1242
Qy	567	MetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe	586
Db	1243	CCCAAGCTCAAGACCGGTGTCGGCTTGTGTCGCTGGTATCGCAGTCGGCGCGCTG	1302
Qy	587	SerPheAspIleAlaTyrGlyHisSerAspLys---LysIleArgTrpHisIleSerLeu	605
Db	1303	CGCTCGACCTGCCAAGCGCTGATGACCAAGGGGCGCATTCGCTGCACCTTTTCCATG	1362
Qy	606	Gly 606	
Db	1363	GGG 1365	
RESULT	10		
LOCUS	AP002568	243184 bp	DNA linear BCT 07-MAR-2001
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 19/20.		
ACCESSION	AP002568	BA000007	
VERSION	AP002568.1	GI:13364484	
KEYWORDS	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)		
SOURCE	DNA.		
ORGANISM	Escherichia coli O157:H7		
REFERENCE	Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae; Escherichia.		
AUTHORS	1 (sites) Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.		
TITLE	Complete nucleotide sequence of the prophage vt2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak		
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)		
MEDLINE	20198780		
REFERENCE	2 (sites) Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T.		
AUTHORS	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MGI655		
TITLE	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		
JOURNAL	20557356		
MEDLINE			
REFERENCE	3 (sites) Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H.		
AUTHORS	Complete nucleotide sequence of the prophage vt1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
TITLE	Gene 258 (1-2), 127-139 (2000)		
JOURNAL	20564182		
MEDLINE			
REFERENCE	4 (sites) Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinagawa,H.		
AUTHORS	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12		
TITLE	DNA Res. 8 (1), 11-22 (2001)		
JOURNAL	21156231		
MEDLINE			
REFERENCE	5 (bases 1 to 243184) Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.		
AUTHORS	Direct Submission		
TITLE	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)		
JOURNAL	genome project.		
COMMENT	Location/Qualifiers		
FEATURES	1. 243184		
SOURCE	/organism="Escherichia coli O157:H7"		
	/strain="O157:H7"		
	/sub_strain="RIMD 0509952"		
	/db_xref="taxon:83334"		
gene	466..2115		
CDS	/gene="ECs5008"		
	466..2115		
	/gene="ECs5008"		
	/note="similar to PGI_ECOLI g11790457 percent identity 100 in 549 aa (Conserved in E.coli K-12)"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="glucosephosphate isomerase"		
	/protein_id="BAB38431.1"		
	/db_xref="GI:13364485"		
	/translation="MKNINPTQAAWALQKHFDEMVDVTIADLFAKGDGRFSFSAT FDDQMLVDYSKNRTTETPLAKLQDLAKGATKSMFSGEKINRTENRAVLHVALR NRSPTILVDGKQVMEPVNAVLEKMTFSEAIIISGEMKGYTGKATDVNIGIGSDI GPYVWTALRPYKNLNMHVSVNDGTHIAEVLKVNDETILFLVASKTPTQTMTN AHSADWFLKAGDERHVAHFALSTNAKVGEGIDTANNFEDWVGGRYSLSWA IGLSIVISIGDNPVELLSGAHMDKHFSTTFAENKLPVLLALIGIWTNNFAGAEIA ILPYDQYMRHFAAYFOCGNMSGKGYVDNRGVVDYGTGPIIWGPGTNGGHAQYQLI HQGTVMPCDFIAPATHNPLSDHOKLLSNFQAQTEALAFKSRVVEQYERDQGD KADTIVYVPKVFEGNRPNTSILLREITPFSICALIAYEHK1PTQGVILNIFTDQW GVELGOLANRLILPELKDDKEISSHSDSTNGLINRYKAWRG"		
gene	2628..2870		
CDS	/gene="ECs5009"		
	2628..2870		
	/gene="ECs5009"		
	/note="similar to YJBE_ECOLI g111790458 percent identity 100 in 80 aa (Conserved in E.coli K-12)"		
	/codon_start=1		
	/evidence=not_experimental		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="BAB38432.1"		
	/db_xref="GI:13364486"		
	/translation="MKKVLGIFALSAALTAATSAWAPVGVGSAAGSAATSVSAGSSA TSVSTVSSAVGVALAATGCGDNGTCTTTTTTSTQ"		
gene	2954..3622		

```
CDS
/gene="ECs5010"
2954..3622
/gene="ECs5010"
/notes="similar to YJBH_ECOLI gll1790459 percent identity
96 in 222 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38433.1"
/db_xref="GI:13364487"
/translation="MYLPLWRVKKRPAPFILICILLQACSAATTKELGNSLWDSLFGT
PGVLQTDQNNPQASYQMLNGQPOLFVLLAFADGQKQWVQDQATLVLTQHGRLV
KTLGLLEVNLAADPLIKPAQIIDAQTMRTMGWTEYQQVRYATARSFVKWDT
GTVKVGSDETAVRVLDSEVSTDOARWHNRWIDSGQIRQSEQLGADYFPVKTLLK
AAKQ"
3619..4356
/gene="ECs5011"
3619..4356
/gene="ECs5011"
/notes="similar to YJBG_ECOLI gll1790460 percent identity
97 in 245 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38434.1"
/db_xref="GI:13364488"
/translation="MIKQITVALLLSVGSASSVFAAGTVKVFSGSSSEAKTLTGAHLI
DLVGQPLRANMPGAVISEELATAAALRQOQALLTRLAEGQADSSADDAANALRQ
QIQAQVLTGROKLNLDQDITVRVAERANPLOGNYTLWGPPTVTLFGLISRGNOP
FTQGRDVASVLSGSLGGADRSYANWYPDGRTOKAPVAYWNRKHVPEHPGSIIVG
LADSVNSEMDNALNLIQTLQRIQ"
4356..6452
/gene="ECs5012"
4356..6452
/gene="ECs5012"
/notes="similar to YJBH_ECOLI gll1790461 percent identity
99 in 698 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38435.1"
/db_xref="GI:13364489"
/translation="MKKRHLISLLALIGSTACYGCTYPAPIGPSQSPGFGVLLQTP
ARMAEGLSLNRYNDQVRYYSASVOLFPWLETTLRVTDVTRQYSSVEAFSGDQY
KDAFDLKLRLWRESYLQVAVGARDIGSTGLDPAEYLIVASKAWGPDFTLIGWGY
LGTSGNKNPLCSADKCYRONSQKAGSIDGSMFHGPASLGGVEYQTPWQPLRL
KLETCGNVQODPAGKLEOKSRFNWCAIYRVDMADYNLSTERANTFMFVILKRN
DLRPSYNDNARPOQPOQDAILOHSVYANQLTLKYNAGLADPOIQAKGDTLVGSE
QVKYTRDREGIVRANRIVMNDLPDGFKTIRVTENLNMPQVTTEDVASLKHILAGEP
LGHETKLAQRKEVPVYPKSTEQGWYIDKSRFDFHIDPVLNSVSGCPENFYMYQVNG
TADLWLDHLLTGSFLFANLANNYDKFTNPPQDHLPRVTRVHVREYVQNDYVNNL
QANTFOHLGNFGYGVVGYLTFMFGAGAEVLYRPLDSNNAFGLDANYVKQDWRSA
KMMKFTDYSVKVTHLTAYWTPSFAODVLVKASGVQYLAGDKGTLETAKRFDSVVV
GGYATINVSKEEYEGEDFTKGVYVSVPLDLFSSGPTSRRAIGTWPLTRDGGQOLGR
KFQILDMSRSRNVNR"
7047..7457
/gene="ECs5013"
7047..7457
/gene="ECs5013"
/notes="similar to YJBA_ECOLI gll1790462 percent identity
98 in 136 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38436.1"
/db_xref="GI:13364490"
/translation="MTSLSRPRVEFISTLIQTVLNLGLLGLILVFLGKGTVHLAD
VLFAPQASKEVLEGLVYVLYFEPFIALIVYFQSGHFPLRYFVYIGITAVRLII
VDHKSPLDLVLYSNAILLVVTWLNCSRLKRE"
```

```
gene
complement(7501..8976)
/gene="ECs5014"
complement(7501..8976)
/gene="ECs5014"
/notes="similar to XYLE_ECOLI gll1790463 percent identity
100 in 491 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="xylose-proton symport"
/protein_id="BAB38437.1"
/db_xref="GI:13364491"
/translation="MNTQYNSYIFSTILVATGLGLFGYDTAVISGTVBSLNTVFVA
PONSLEAANSLGFCVASALIGCIIGALGCGYCSNRFGRDLSLKIAVLPFFISGVGS
AWPLGFTSNPNTVPVYLVAGVYFPEVYIRIIGIGVGLASMLSPMYIARLAPAHIR
KWLFSNQPAIIFGQLLYVCVNYFIARSGDASMLNTDGMRYMFASECIPALLFLMLLY
TVPESPRLMSRKQQAEGILKIKNGNTLATQAVQBEIKHSLDHGRKRTGGRLLMFGVG
VIVIGVMSLFPQFVGVNLYIYAPFVKTLGASTDIALLOTIIVGVINLTFTVLAIM
TVDFGRKPLQIIGALGMAIGMFSUGTAFYQAPGIVALLSNLKYVFAAFAMSGPVCW
VLLSEIFPNAIRKALAIYAAQWLANFYVSWTFPMMDKNSLWVAHFHNGFSYWIYGC
MGVLAALFMKRFVPETKGKTLBELEALWEPETKKTQQTATL"
complement(9348..10238)
/gene="ECs5015"
complement(9348..10238)
/gene="ECs5015"

Alignment Scores:
Pred. No.: 5.65e-14 Length: 243184
Score: 380.50 Matches: 145
Percent Similarity: 39.46% Conservative: 87
Best Local Similarity: 24.66% Mismatches: 245
Query Match: 11.99% Indels: 111
DB: 1 Gaps: 20

US-09-857-669-2 (1-609) x AP002568 (1-243184)
QY 74 Leu1eThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu 93
||| ||| : : : : : ||| : : : : :
Db 211436 CTTTCTAGATTGAAAGTGATGAGTGCAGCCAGACCGTCGC-----TTTCGCGCA 211486

QY 94 GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerLysVal 113
||| : : : : : ||| : : : : : ||| : : : : :
Db 211487 CGCGTCGATGATGCTATCCGCGAAGGCTGAAACGCGTGGGTATTACCAGCGGCACATT 211546

QY 114 SerLeu-----ThrGluLysAspGlyAlaTyrThrVal-----HisIleThr 127
||| ||| ||| ||| ||| : : : : :
Db 211547 GAATTGTGATCCGTCACCCGCCAAAGAGGCGGCGAGTATTGTCGCCAAGTCACG 211606

QY 128 ProGlyProArgThryIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSer 147
||||| ||| ||| ||| : : : : :
Db 211607 CCAGGCGTCCCGGTGTTAATTTGGCGGCACCGCATGTGTATTCGCGGGGCGCGGACG 211666

QY 148 AspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGly 167
||| : : : : : ||| : : : : : ||| : : : : :
Db 211667 GAT-----AAAGACTATTTTGAAA-----TTGCTCGATACTCGCCCGGCTATTGC 211711

QY 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
: : : : : : : : : : : ||| ||| : : : : :
Db 211712 ACGGTACTGAACCCAGGCGGATATGAAAAATTTCAAAGCTTCAACAGCATGCGGTG 211771

QY 188 ArgLysGlyTyr-----ProLeuAlaLysLeuGlyAsnThrArgAlaAla 202
||||| ||| ||| ||| ||| : : : : :
Db 211772 CGTAAAGTTATTTCGATAGCGAATTTACCAAGCGCAGCTGGGC-----ATTGCG 211822

QY 203 ValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIle 222
||| ||| : : : : : ||| : : : : :
Db 211823 CTCGGCGCTGCATAAAGCCCTCTGGGATATTGAT-----TATACAGTCGCGCACTTAC 211876

QY 223 AlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSerGly 242
||||| ||| ||| ||| : : : : :
Db 211877 CGCTTTGGCGATGTGACCTTTGAAGGATCACAA---ATCCGCGATGAATACCTGCAAAAT 211933

QY 243 LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGln 262
```


sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

Location/Qualifiers

1. 10819

/organism="Escherichia coli K12"

/strain="K12"

/sub_strain="MG1655"

/db_xref="taxon:83333"

36. 36

/note="central position to predicted promoter: -73.5"

/bound_moiety="flv predicted site"

86. 114

/note="factor Sigma70; predicted +1 start at 4435250"

156. 656

/gene="ytfI"

/note="b4215"

156. 656

/gene="ytfI"

/function="orf; Unknown"

/note="ol66"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC77172.1"

/db_xref="GI:1790661"

/translation="MLPRIRHNNF IGAVELFVKSYTKTHSNFNNHFAFKKDWI
SNFSLTLREFRCATQIDKSGYQVLSSKNETVHAMDKFLISFLKDNGAETMTLR
GSGFEVEIPITINEYNSPMDFKREFPLEQTDCCMPGIFYRKNSLYRKGKLVIFTR
RLVTCL"

complement(1156..1710)

/gene="ytfI"

/note="b4216"

complement(1156..1710)

/gene="ytfJ"

/function="orf; Unknown"

/note="f184; 100 pct identical amino acid sequence and

equal length to ytfJ_ECOLI SW: P39187"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC77173.1"

/db_xref="GI:1790662"

/translation="MTLRKILALCTLLPMMASAHQFTGQRPPIGITDGEIVLDK
DQSYTWSAQVGVKRVQLHAGTSAKKNATLIEAKSAPLPHDPTQTTIVT
DDALPGSGMFRVSRSSKSLKLYPWSQFTVDSNGVALGAWQDESSAVVLDKGRVQ
WAKDGLTPREVQVMDLQKLLK"

complement(1790..1807)

/note="factor Sigma54; predicted +1 start at 4436912"

complement(1882..1910)

/note="factor Sigma70; predicted +1 start at 4437004"

1924..1951

/note="factor Sigma70; predicted +1 start at 4437087"

1996..2241

/gene="ytfK"

/note="b4217"

1996..2241

/gene="ytfK"

/function="orf; Unknown"

/note="o81; 100 pct identical amino acid sequence and

equal length to ytfK_ECOLI SW: P39318"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC77174.1"

/db_xref="GI:1790663"

/translation="MAPHRNTQOETMKIFORYNPLOQVAKYKILFRGLYIKDVG
FEDKGLIPKVKDLHLSVMSEVNRQVMRLQTEMA"

complement(2320..3663)

/gene="ytfL"

/note="b4218"

CDS

complement(2320..3663)

/gene="ytfL"

/function="putative transport; Not classified"

/note="f447; 100 pct identical amino acid sequence and

equal length to ytfL_ECOLI SW: P39319"

/codon_start=1

/transl_table=11

/product="putative transport protein"

/protein_id="AAC77175.1"

/db_xref="GI:1790664"

/translation="MLNSILVILCLJAVSAFFSMSEISLAASRKIKLKLADENGINA
QVLNMQENGFVTVQIGLNAVATLGGIVGDAAFSPAFHSLSFRYSMSAELSQLSF
ILFSFLVTGMFLLADLTTPKRICGMIAPEALVRIINPMRFLYVCTPLVWFFNGLAHI
IFRFKLPVVRKDDITSDDIYAVVAGALAGVLRKOEHLIENVFELSRVPSMPP
LNSVIFDLHEDEQSILKNKVAHPHSKFLVCNEDIDHIIIGYVDSKDLLNRVLNQSLA
DLVQGLREQIIVARDENSLIDGGTPIDDVMRVLOIDEPGSGNYETIGGFMMFLKK
IPKRTDSVPAGYKEVDIDNFRIDQLLVLRIDSKATALSPLKLPDADKKEESA"

complement(3722..3750)

promoter

/note="factor Sigma70; predicted +1 start at 4438844"

complement(3785..3813)

promoter

/note="factor Sigma70; predicted +1 start at 4438907"

3905..3938

repeat_region

/note="REP (repetitive extragenic palindromic) element;

contains 1 REP sequence"

complement(3986..4624)

/gene="msrA"

/note="b4219"

complement(3986..4624)

/gene="msrA"

/function="enzyme; Proteins - translation and

modification"

/note="f212; 100 pct identical amino acid sequence and

equal length to PMSR_ECOLI SW: P27110"

/codon_start=1

/transl_table=11

/product="peptide methionine sulfoxide reductase"

/protein_id="AAC77176.1"

/db_xref="GI:1790665"

/translation="MSLFDKKHLVSPADALPGRTNTPMPVATLHVNGHSTNVPDGM
EIAFAMGCFWGVRELFWQLPCYVSTAAGYGGYTNPFTYRVCSCDGHAAVRLIYD
PSVISQQLQVFNWHDPAQMRQNDHGTQYRSATYTPQDAARAASLERFQAA
MLAADDHRTTETIANATPFYAEDDHQYLLHKNPYGGIGGIGVCVLPPEA"

4662..4690

promoter

/note="factor Sigma70; predicted +1 start at 4439826"

complement(4687..4718)

promoter

/note="factor Sigma70; predicted +1 start at 4439809"

4751..4762

protein_bind

/note="central position to predicted promoter:61"

/bound_moiety="AraC Predicted site"

4830..6563

/gene="ytfM"

/note="b4220"

4830..6563

/gene="ytfM"

/function="orf; Unknown"

/note="o577; 100 pct identical amino acid sequence and

equal length to ytfM_ECOLI SW: P39320"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC77177.1"

/db_xref="GI:1790666"

/translation="MRYIROLCCVSLCLSSAAVNVRLQVGLSGOLEKNVRAOLS
TIESDEVTPDRRFRASVDDAIREGLKALCYOPTIEFDLPKPKGRQVLIKVTG
PVLIGGTDVVLRGARTDKDYLLDTRPAITVLNQGDYENFKKSLTSIALRKGIFD

Alignment Scores:

Pred. No.:

1.11e-15

Length:

Matches:

10819

379.50

145

Score:

Conservative:

87

Percent Similarity:

39.46%

Best Local Similarity:

24.66%

Mismatches:

245

Indels:

111

Query Match:

11.96%

DB:	1	Gaps:	20
US-09-857-669-2 (1-609) x AE000493 (1-10819)			
Qy	74	LeuIleThrGlnGlnGlnGluValLeuAspLysGluInThrGlyPheLeuAlaGlu	93
Db	4956	CTTCTACGATTGAAGTCAATCAAGTCAGCGCAGACCGTCGC-----TTTCGCGCA	5006
Qy	94	GLUALaprasnValLysThrMetLeuargSerLysGlyTyrPheSerSerLysVal	113
Db	5007	CGGCTCGATGATCCATCGCGAAGGTCTGAAGCGCTGGGTTATTACCAAGCGACCAT	5066
Qy	114	SerLeu-----ThrGluLysAspGlyAlaTyrThrVal-----HisIleThr	127
Db	5067	GAATTTGATCTCGCTCCCGCAAAAGAGCGCGCAGGTATTGATGTCGCAAGTCACG	5126
Qy	128	ProGlyProArgThrLysIleAlaasnValGlyValAlaIleLeuGlyAspIleLeuSer	147
Db	5127	CCAGGCGTCGCGGTGTAATTGGCGCACCGCATGTGGTATTGCGCGCGCGCGCGGAC	5186
Qy	148	AspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnInProValGly	167
Db	5187	GAT-----AAAGACTATTTGAAA-----TTGCTCGATACTCGCCGGCTATTGCG	5231
Qy	168	SerAspPheAspClnAspSerTrpGluasnSerLysThrServalLeuGlyAlaValThr	187
Db	5232	ACGGTACTGAACAGCGGCGATTATGAAATTTCAAAAAGTCCTTAACCGACATGTCGTTG	5291
Qy	188	ArgLysGlyTyr-----ProLeuAlaLysLeuGlyAsnThrArgAlaAla	202
Db	5292	CGTAAAGTTATTTCGATAGCGCAATTTACCAAGCGCAGCTGGGC-----ATTGCG	5342
Qy	203	ValasnProAspThrAlaThrValAspLeuasnValValAspSerGlyArgProIle	222
Db	5343	CTGGCGTCGATAAAGCCTCTGGGATATTGAT-----TATACAGTGGCGAACGTTAC	5396
Qy	223	AlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGly	242
Db	5397	CGGTTTGGGCATCTGACCTTTCAAGGATCACAA---ATCCGCGATGAATACCTCGA	5453
Qy	243	LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGln	262
Db	5454	CTGGTCCGCTTAAAGAGCGCATCAGTACGANTCGAAGATCTGGCAGAACTGAACCCG	5513
Qy	263	AlaLeuGluClnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu	282
Db	5514	CGACTTTCTGCTACCGGCTGGTTAACTCGGTGGTGGTCCACCAATTTGTATAAGCG	5573
Qy	283	GlnGlyAspArgVal---ProValLysValSerValThrGluValLysArgHisLysLeu	301
Db	5574	CGCAAAAGCAAGTATTACCATTTAGCGGCGTGGTTCCGCGCGCAACAGAAACACCATC	5633
Qy	302	GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr	321
Db	5634	GAACCGGGTCGGTTACTCTACGACGTGGGACCGCGCTGGAAGCAGCAGTGG-----	5687
Qy	322	TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu	341
Db	5688	-----AAAAGCGTGGATGAACCTTAT-----GGTCACAGT	5720
Qy	342	ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer	361
Db	5721	CTGACCACCAAGTACTAGTATTTCCGCGCGCAACAG-----ACCCCT	5762
Qy	362	AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly	381
Db	5763	GACTTCAGCTATAAATGCCGCTGCTGAAGAT-----	5795
Qy	382	IleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAla	401
Db	5796	-----CCACTGGAAACAATATTTTGGT	5819
Qy	402	GluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu	421

Db	5920	CAGGCG	---GGTTTTTAAGCGCACTGACGCTGAACGATACCGAATCTGACTCCACTCAGCTG	5876
Qy	422	ThrAlaSer	-----TrpLysArgGlnLeuLeuAsnVal	433
Db	5877	GTGGCTTC	TGCTGCTACTCGGATCTCTCCAGCGGCTGGCAGCGTCCCATTT	5927
Qy	434	LeuHisProGluAsnGlyHis	TrpLeuAspGlyLysIleGlyThrThrLeuGlyThrPhe	453
Db	5928	CTCGCGTGGAGTCTCGACACATTT	CTACTACTCGGGTGAATATACCAATACCAACGATGCTGTTT	5987
Qy	454	LeuSerSerThrAlaLeuIle	ArgThrSerAlaArgAlaGlyThrPhe	469
Db	5988	TATCCTGGGTGATGATTAGCGGCACGCGTTCT	CTCGTGGCTGATGCCAACCTGGGCG	6047
Qy	469	-----	-----	469
Db	6048	GACTCGCAACGCTACTCTAT	CGACTACTCCAAACAGCGCTTCCAGATCGCATTTTC	6107
Qy	470	-----	-----PheThrProGluAsnLysLysLeuGlyThr	481
Db	6108	TCCGTTTTTC	CAGCGCAGACGCTCTGGATCCGACACTGACGATCCCATCTTTTGT	6167
Qy	482	IleArgGlyClnAlaGly	ThrValAlaIleArgAspAsnAlaAspValProSerGlyLeu	501
Db	6168	ACAGCGGGCAGCGTGGCTGGATG	GAACCGGTGATTTCCACAAGTACCGCGGATCG	6227
Qy	502	MetPheArgSerGlyAlaSer	SerValArgGlyThrGluLeuAspSerIleGlyLeu	521
Db	6228	CGTTCTTC	CGCGGGGCGACGCGATATTCTGGTGGCTACAAATACAAATCTATCGCTCG	6287
Qy	522	AlaGlyProAsnGlySerVal	LeuProGluArgAlaLeuLeuValGlySerLeuGlyThr	541
Db	6288	AATACGCCAACGCGTCACTG	AAAGCGGCTCGAAGTTGATACCGGATCGCTGGAAATAC	6347
Qy	542	GlnLeuProPheThrArgThr	LeuSerGlyAlaValPheHisAspMetGlyAspAlaLeu	561
Db	6348	CAGTACAAACGTACCGGAAATGT	GTGGGCGCGGTGTTCTCGATAGTGGCAGACGGA	6407
Qy	562	AlaAspPheLysArgMetLys	LeuLysHisGlySerGlyLeuGlyValArgThrPheSer	581
Db	6408	AGGATATTCCGCGCAGCAGCACTT	AAACCGGTACCGGCTCGCGTGGCGTGGGAATCG	6467
Qy	582	ProLeuAlaProPheSerPhe	AspIleAlaThrGlyHisSerAspLys	598
Db	6468	CCGTCGGGCGCAATCAAACTCG	ATTTTCCGTACCGTCCGCGATAAAGAACGACGCGG	6527
Qy	599	IleArgThrHisIleSerLeu	Gly 606	
Db	6528	TTACAGTTTACATCGGCTGGGG	6551	
RESULT 12	AE005654	LOCUS	Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 273 of 290.	
ACCESSION	AE005654	VERSION	AE005654.1	GI:12519212
KEYWORDS	Escherichia coli O157:H7 EDL933.	SOURCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	
REFERENCE	1	(bases 1 to 11760)		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamous,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.			
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7			
JOURNAL	Nature 409 (6819), 529-533 (2001)			
MEDLINE	21074935			

PUBMED 11206551
 REFERENCE 2 (bases 1 to 11760)
 AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 source Location/Qualifiers
 1. 11760
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 126..866
 /gene="cysQ"
 /note="Z5825"
 126..866
 /gene="cysQ"
 /function="phenotype: Central intermediary metabolism: Sulfur metabolism"
 /note="Residues 1 to 246 of 246 are 99.18 pct identical to residues 1 to 246 of 246 from Escherichia coli K-12 Strain MG1655: B4214"
 /codon_start=1
 /transl_table=11
 /product="affects pool of 3'-phosphoadenosine-5'-phosphosulfate in pathway of sulfate synthesis"
 /protein_id="AAG59412.1"
 /db_xref="GI:12519213"
 /translation="MLDQVCQARNAQDAIMQVVDGAKPMQVSKANSPTVAADIAA HTVMGLRFLTPDIPVLSEEDPGWEVRQWRQVWLDVLDGKTEIKRNGEFTVAI ALIDHGPILGVVYPMNVY SAAGKAWKECCVQKIQVRDARPLVIVSRSHAD AEIKYLOQHQHTTSGISLKFCGLVAEGQAQLYPRFGPTNIWDTAAGHAVAAGA HVIDWQCKPLDYPRESFLNPGFRVSIY"
 1078..2016
 /gene="Z5826"
 1078..2016
 /gene="Z5826"
 /function="orf; Unknown function"
 /note="Residues 1 to 133 of 312 are 98.49 pct identical to residues 1 to 133 of 166 from Escherichia coli K-12 Strain MG1655: B4215"
 /codon_start=1
 /transl_table=11
 /protein_id="AAG59413.1"
 /db_xref="GI:12519214"
 /translation="MLPRLHNNFTGAVELFVKSSYTKTHSNFNNHIAFKKKDWI SNYDSLTLREFRCATQIDKSYQVLSKNETVAMDKFLISLKDNGAEYTMTLR GSFEYEIPITITNEYNSFMDFKNREFPLEQNRRLYAMDILQKQSDIPKRIKGIHQ AIGDVSGLYALDDIIVSKLRGRKFELQPGGIGKOCGWYIYKIIDNFAIVIESLG FALKIYGGDERFRNGSSVLEDEDYSLIVNVLVAGCQOVELAQVDAIVSANLAADS DTNKEKICEYKYSTIEAFKKEQLALPVLVRRKNSET"
 complement(2079..2633)
 /gene="ytfJ"
 /note="Z5827"
 complement(2079..2633)
 /gene="ytfJ"
 /function="orf; Unknown function"
 /note="Residues 1 to 184 of 184 are 98.91 pct identical to residues 1 to 184 of 184 from Escherichia coli K-12 Strain MG1655: B4216"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG59414.1"
 /db_xref="GI:12519215"
 /translation="MTLRKILATCLLLPMMASAHQETGQRPVPPIGITDRGELVLDK

DOFSYKTWNSAOLVGVKRVLOHIAGTSAKEKNATLIEAIKSAKLPHDRYOTFTIVNT
 DDAIPGSGMFVRSSLESNNKLYPWSQFIVDSNGVARGAQLDESSAVVLDKRGRO
 WAKDGAUTQEEOVQVMDLLQKLLK"
 2919..3164
 /gene="ytfK"
 /note="Z5828"
 2919..3164
 /gene="ytfK"
 /function="orf; Unknown function"
 /note="Residues 1 to 81 of 81 are 100.00 pct identical to residues 1 to 81 of 81 from Escherichia coli K-12 Strain MG1655: B4217"
 /codon_start=1
 /transl_table=11
 /protein_id="AAG59415.1"
 /db_xref="GI:12519216"
 /translation="MAPHRNTQOEITMKIFORYNPLOVAKYVYLFRGLYIKDVCA
 FEEDKGLIPKVKDKLHLSVMSEVNRQVRLQTEMA"
 complement(3260..4603)
 /gene="ytfL"
 /note="Z5829"
 complement(3260..4603)
 /gene="ytfL"
 /function="putative transport; Not classified"
 /note="Residues 1 to 447 of 447 are 100.00 pct identical to residues 1 to 447 of 447 from Escherichia coli K-12 Strain MG1655: B4218"
 /codon_start=1
 /transl_table=11
 /product="putative transport protein"
 /protein_id="AAG59416.1"
 /db_xref="GI:12519217"
 /translation="MLNSILVLCILAVSAPFMSFSEISLAASRKIKLKLAEQGNINA
 ORVLNMGENPGMEFTVVQIGLNAVALIGSIVGDAAFSPAFSLSRYSNAELSQLSF
 ILSFLVIGFILFADLTPRRIGMAIPEAVALRIINPMRFCLVCTVLPWFNGLANI
 IRRFKLPVMYKDDITSDDIYAVEAGALAGVLRKOEHLIENVFELSRTVPSSMTP
 RENYVFDLHREKDESLKNKVAEPHPSKFLVCNEDIDHIGYVDSKDLNRYLANOSLA
 LNSGQIRNTLIVPDTLTSEALESPKTAGEFAVIMNEYALVWGITLNDVMTLNG
 DLVGQGEQIVARDENSLIDGCTPIDVMRVLDIDEPQSGNYETIGGPMFMRLRK
 IPRKTDVSKAGYKFEVVDIDNYRIDQLLVTRIDSKATALSPLKPAKDKEESA"
 complement(4926..5564)
 /gene="msrA"
 /note="Z5830"
 complement(4926..5564)
 /gene="msrA"
 /function="enzyme: Macromolecule synthesis, modification: Proteins - translation and modification"
 /note="Residues 1 to 212 of 212 are 99.52 pct identical to residues 1 to 212 of 212 from Escherichia coli K-12 Strain MG1655: B4219"
 /codon_start=1
 /transl_table=11
 /product="peptide methionine sulfoxide reductase"
 /protein_id="AAG59417.1"
 /db_xref="GI:12519218"
 /translation="MSLFQKHLVSPADALPGRNTPMPVATLHVNGHSMTPVPGME
 IATFAMCGFWGVERLFWLGHVYSTAGYTGTYTPNTRYEVCSGDTGHAARVIVYD
 PSVTSYQLLOLVENEDHPAGMROGNDHGTOYRSNLYPLPEODAAARASLERFQAA
 MLAADDRHITTEIANNTFYAEDDHQYLLKKNPYGCGIGGICVCLPPEA"
 5770..7503
 /gene="ytfM"
 /note="Z5831"
 5770..7503
 /gene="ytfM"
 /function="orf; Unknown function"
 /note="Residues 1 to 577 of 577 are 100.00 pct identical to residues 1 to 577 of 577 from Escherichia coli K-12 Strain MG1655: B4220"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG59418.1"
 /db_xref="GI:12519219"

```
/translation="MRVIRQLCCVSLCLSGSAVANPELOVEGXSQLEKNVRAQLS
TLESDEYTPRRFARVDDAI REGLNALGYQTFEPLRPPPKGRGVLI AKVTG
PVLIGGTDVILRGADKDYILKLLDTRPAIGTVLNQGDYENFKSLYSIALAKVTG
SEPTKAQGLALGHLKAFKADIDYNSGERYFEGHVFEGSQINDEYLNQLVFPKGE
ESKOLAEINRLSATGSFNSVVAPOFDRARETKVLPVTGVVSPRTENTIEYGVYST
DVGPRVATWKKKPMNSYSHSLTSTSISAPQTLDFSYKMPLLKNPLEQYLYVQGG
KRTLDNEDSTTLVASRYWDLSSGWRAINRLWSDLHFTOGELTNMTMLFVPCVM
SRTSRGLMPTMGDSORYSIDYNTAGMSDVDFSVFOAONVMI RTLDYDRHFRVTR
LGMLETGDFKVPDPLREFFAGDERSIRGYKSIAPKYANGDLKASKLINGSLEYQ
NYTKWCAVAVDSGEAVDIRSRDFKTGTGCVGRWESPVGPIKLDFAVPVADNDEH
LQFIIGLPEL"
7500..11279
/gene="ytnf"
/ncbi="25832"
7500..11279
/gene="ytnf"
/function="orf; unknown function"
/notes="Residues 1 to 1259 of 1259 are 98.80 pct identical
to residues 1 to 1259 of 1259 from Escherichia coli K-12
Strain MG1655: B4221"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/db_xref="GI:12519220"
/translation="MSLWKKISIGVVIVILLIGSVAFVLGTTSGHLHVFKAARWVP
GLDQKVTGMRDUTLSDRVEQGVAVRAGNLHLAVGLECLWNSVSCINDLALDKDIO
VNIQSKMPSQVEEEDSGPDLSTPYITLIRVALDNVNIKIDDTTVSMVDFTS
LNWQKLTUPLKSLKGLLALPKVAEVAQEEVPEKIENPOEEKXXGETLKOLFSP
PVLPEMTDHLPLNLNTEFEFQQLRVGTGDTITVTRMLLKVSSIDGNTKDLALDIDS
NOGXNASTGAQLSDNMPVDITLNSLTINVEPLKGRVKLVKGALREOLEIGVNLSGP
VMDLRAQTLAEAGLPLNVEVNSKOLYMPFTGEKYOQADDLKLTCKMTDITLSMR
TAVKLEIPPATITLDAKNGEQVNLDTVAALSGKTELKALLDMQOAIISWRGELTL
NGINTAEIPWPSKLNCLIKLTRGSLYGGTQWQVEPELKLTVNWKVKNVYDGLKGN
```

gene

CDS

Alignment Scores:

```
Pred. No.: 1.24e-15 Length: 11760
Score: 379.50 Matches: 158
Percent Similarity: 38.88% Conservative: 92
Best Local Similarity: 24.57% Mismatches: 259
Query Match: 11.96% Indels: 135
DB: 1 Gaps: 22

US-09-857-669-2 (1-609) x AE005654 (1-11760)

Qy 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
Dy 5801 GCTTACTCTGCTTAACGGGATCTG----- 5824
Qy 39 AsnLysSerProAspThrCluserValLysLeuLysProLysPheProValArgIleAsp 58
Dy 5825 ---CCGTCCCGCGAAGCTCCGCTACAGGTCGAGG-----GGT 5860
Qy 59 ThrGluAspSerGluIleLysAspMetValGluGluLisLeuProLeuIleThrGlnGln 78
Dy 5861 YATCGGCACAGCTGGAAGAAGACGTTCTGCGCAGCTTCTACGATTG----- 5908
Qy 79 GluGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsn 98
Dy 5909 -----AAGAGTATGAGTGACCCGACAGCTCGCTTCGCGCAGCGGTCCGAT-GATGCT 5961
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu----- 115
Dy 5962 ATCCGGGAAGGTGTGAAGCGGTGTATTATACGACGCGACCATTTGAATTTGATCTCCGT 6021
Qy 116 ---ThrGluLysAspGlyAlaTyrThrVal-----HisIleThrProGlyProArgThr 132
Dy 6022 CCACCCCAAGAAAGGCGGCGATGATTATGCCAAAGTCAAGCCGCGCGCGGTG 6081
Qy 133 LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Dy 6082 TTAATTGGCGGACCGCATGTGTATTTCGCGGGGGCGCGGACCGAT-----AAA 6132
```

```

Db 7063 CAGACGGTCTGCATCCGACACTGTACGATCGCATCGTTTGTGTGACACCGCGCAGGTG 7122
      :::::      :::      |||      :::::      |||||
Qy 487 GtYtThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
      :::::      |||      |||||      |||      |||      :::::
Db 7123 GGCTGGATTGAACCGGTGATTTTCGACAAAGTACCGCGGATCGCTGTTTCTTCGCGGG 7182
      :::::      |||      |||||      |||      |||      :::::
Qy 507 GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGly 526
      |||      :::::      |||||      :::::      |||||      |||
Db 7183 GCGGATCGCAGTATTCGGGGTACAAATACAAATCATATCCGCCGAAATACCGTACCGT 7242
      |||      :::::      |||||      :::::      |||||      |||
Qy 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThr 546
      :::::      |||||      |||||      |||||      |||||      |||
Db 7243 GACCTGAAAGGGGCTCGAAGTTGATACCGGATCGTGGAAATACCAACGTCAGCC 7302
      |||      :::::      |||||      :::::      |||||      |||
Qy 547 ArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArg 566
      |||||      |||||      |||      |||||      |||||      |||
Db 7303 GGAATGTTGGGGCGCGTGTTCGATAGTGGCAGCGGTAGCGATATTCGCGCG 7362
      |||||      |||||      |||      |||||      |||||      |||
Qy 567 MetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe 586
      |||      |||||      |||||      |||||      |||||      |||
Db 7363 AGCGACTTTTAAACCGGTACCGGGTGGCGGTACGCTGGGAATCGCGGTGGCGCAATC 7422
      |||      |||||      |||||      |||||      |||||      |||
Qy 587 SerPheAspIleAlaTyrGlyHisSerAspLysLys-----IleArgTrpHisIle 603
      |||      |||      :::::      |||||      |||||      |||
Db 7423 AACTCGATTTTGGCTACCGTGGCGGATAAAGACGACACCGGTTACAGTTTACATC 7482
      |||      |||      :::::      |||||      |||||      |||
Qy 604 SerLeuGly 606
      |||||
Db 7483 GGTCTGGG 7491

RESULT 13
ECOW93
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL

COMMENT

FEATURES
source
1..338534
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/map="92.8 to 100 minutes"
/note="This sequence comprises the following lambda

```

```

clones: DD928(EC27-278), DD930(EC30-188),
DD933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070),
DD941(EC17-142), DD945(EC17-8), DD947(EC24A-34),
DD949(EC22-169), DD952(EC27-297), DD953(EC27-409),
DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2),
DD962(EC21-104), DD965(EC23A-40), DD968(EC30K660A-4pp),
DD970(EC19-202), DD974(EC19-61), DD975(EC18-233),
DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3),
DD984(EC17-101), DD987(EC18-115), DD990(EC17-136),
DD992(EC18-282), DD995(EC17-115), DD997(EC18-113); M13mp19
or Janus vectors were used for subcloning".
1..1885
/note="1885 bp overlap with end of GenBank Accession
Number U00006 (ECOWU89)"
complement(<1..398)
/note="ORF_f326 of GenBank Accession Number U00006
(ECOWU89)"
/codon_start=1
/transl_table=11
/label_ORF_f326p
/protein_id="AAA96985.1"
/db_xref="GI:1263173"
/translation="MGFTTRVKSEASEKPFNFALFMDKYGTFFILAIIVAFGSLSP
EYFTTNNITQIFVQSSVTVLIGMGFEFFAILVAGIDLSVGAILUSGMVTAKLMLAGV
DPFLAAMIGVLYGGALGALGNVNTGL"
277..397
/note="predicted bend of 75.10 degrees"
complement(377..1909)
/note="ORF_f510"
/codon_start=1
/transl_table=11
/label_ORF_f510
/protein_id="AAA96986.1"
/db_xref="GI:536931"
/translation="MATPYISMAGIKSGFPHALKSVNLTVYPGEIHALLGENGAGK
STLMKVSIGHEPTKGTITINNYSKNDLHAAOLGIGIYOEUSLDELTVLENLY
IGRHLTKKICGVNIIDRENRVRAMLLRVLGKLDLDEKVAANLSISKOMLEIAKTL
MLDAKVIIMDEPTSSLTNKEVDYFLIMNQLRKEGTAVIYISHKLAETRRICDRYTM
KDGSSVCVGIVSDVNDIIVRLMVGRELQNRNAMKNVSNLAHETVEFRNVTSRDR
KVKRIMAYITPESRPDGFPPNFSTIAONMAISPSLKGCKYKGMGLFHEVDQRTAENC
AVKGMAYITPESRPDGFPPNFSTIAONMAISPSLKGCKYKGMGLFHEVDQRTAENC
RELLAKLCHSVNQNTLSESGNQKVLISKWLCGCCPEVIFDEPRGIDVGAKAEIYK
VMRLQADDGKVLMSVSELPEITVCDRIAFVCEGRLTQILTNRDDMSEEEIMAWALP
QE"
809..931
/note="predicted bend of 74.75 degrees"
1768..1889
/note="predicted bend of 75.87 degrees"
complement(1983..2010)
/note="promoter matrix score of 55; putative; resembles
filh-regulated promoter"
/standard_name="REP; repetitive extragenic palindromic
element"
/note="contains 1 REP sequence"
complement(2036..2971)
/note="ORF_f311"
/codon_start=1
/transl_table=11
/label_ORF_f311
/protein_id="AAA96987.1"
/db_xref="GI:536932"
/translation="MNKLYKFSGTGLVGLMLSTSAFAAAEYAVVLKTLSPNFWDMKK
GIEDEAKTLGVSDVDFASPSGEGFQSOQLFEDLSNKNYKGIATAPSLDSVNLVMPVAR
AWKKGILVNLDEKIDMDNLKAGGVEAFVTIDNVAVGAKGASFLIDKLGAEGEVA
TIEGKAGNASCGEARRNGATEAPFKASQIKLYASQPADNDRIKALOVATYRQIRPNIK
ALIYCANDTMANGVQAQAVNAGKTGVLVGTGDIPEARAKWVCAQGTATVQAQNPADIG
ATGLKLMVDAAEKSGKVIPLDKAPEFLVDVLSILVTQ"
3010..3121
/note="predicted bend of 73.22 degrees"
complement(3030..3953)
/note="ORF_f307"
/codon_start=1

```



```
QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyrAspTyr 321
Db 134017 GNAACGGGGTGGTACTACGAGCGTGGGACCGCGCTGAACGACGCTGG----- 134070
QY 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
Db 134071 -----AAAAGCGCTGGATGAACCTCTAT-----GGTCACAGT 134103
QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 134104 CTGACCACCACTACTAGTATTTCCGCCCGGAACAG-----ACCTC 134145
QY 362 AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGlyLysArgAlaPheSerGlyGly 381
Db 134146 GACTTCAGCTATAAATCCGCTGCTCAAGAT----- 134178
QY 382 IleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyValaGluPheLeuAla 401
Db 134179 -----CCTCGACATATATTTGGTG 134202
QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu 421
Db 134203 CAGGCG---GGTTTAAAGCAGCTGACCTGAACGATACCAGAACTGACTCCACTACGCTG 134259
QY 422 ThrAlaSer-----TrpLysArgGluLeuAsnAsnVal 433
Db 134260 GTGGGTTCTCCTACTGGGATCTCTCCAGCGGCTGGCAGCGTGCCATT-----AAC 134310
QY 434 LeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPhe 453
Db 134311 CTGCGCTGGAGTCTCGACCACTTTACTCAGGTTGAATATACCAATACCAGATGCTGTT 134370
QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyThrPhe----- 469
Db 134371 TATCTGGGGTGATGANTAGCCGACCGCTTCGTGGTGGCTGATGCCAACCTGGGGC 134430
QY 469 ----- 469
Db 134431 GACTCGCAACGCTACTATCGACTACTCAACACGGCCTGGGTTTCAGATGTCGATTTC 134490
QY 470 -----PheThrProGluAsnLysLysLeuGlyThr-----PheIle 481
Db 134491 TCCGTTTTCAGGCGCAGACGCTGGATCCGACACTGTAGATCGCCATCGTTTGT 134550
QY 482 IleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeu 501
Db 134551 ACACCGGCAGCGCTGGCGTGGATTGAACCGGTGATTTCGACAAAGTACCGCCGATCTG 134610
QY 502 MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu 521
Db 134611 CGTTCTTCGCGGGGGGCGGACCGCAGTATTCGTGGCTACATAACAAATCTATCCTCCG 134670
QY 522 AlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValcylSerLeuGluTyr 541
Db 134671 AAATACGCCAACGCTGACCTGAAAGGGCGCTCGAAGTTGTATACCGGATCGCTGGAATAC 134730
QY 542 GlnLeuProPheThrArgThrLeuSerGlyValaValPheHisAspMetGlyAspAlaAla 561
Db 134731 CAGTACACGTCGACCGGAAATGCTGGGCGCGCTGTTTGTCTGATAGTGGCGAACGGTA 134790
QY 562 AlaAsnPheLysArgmetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 134791 ACGGATATTCGCCGACGAGCTTTAAACACGGTACCAGGGGTGGCGTGGCTGGGAATCG 134850
QY 582 ProLeuAlaProPheSerPhePheAlaTyrGlyHisSerAspLysLys----- 598
Db 134851 CCGGTGGGCGCCAACTCAACTCGATTTGGCGTCCCGGTACCGGATAAACACGACCGG 134910
QY 599 IleArgTrpHisIleSerLeuGly 606
Db 134911 TTACAGTTTTATACGTCGTGGG 134934
RESULT 14
```

```
AX033468
LOCUS AX033468 1737 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 1 from Patent WO0047737.
ACCESSION AX033468
VERSION AX033468.1 GI:10280229
KEYWORDS Haemophilus influenzae.
SOURCE Haemophilus influenzae.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
AUTHORS Ruelle,J.L. and Thonnard,J.
TITLE Haemophilus influenzae rd outer membrane sequences used as vaccine
JOURNAL Patent: WO 0047737-A 1 17-AUG-2000;
RUELLE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ; THONNARD
JOBELLE (BE)
FEATURES
source Location/Qualifiers
1..1737
BASE COUNT 517 a 309 c 380 g 531 t
ORIGIN
/organism="Haemophilus influenzae"
/db_xref="taxon:727"
Alignment Scores:
Pred. No.: 2,36e-16 Length: 1737
Score: 374.00 Matches: 153
Percent Similarity: 36.88% Conservative: 93
Best Local Similarity: 22.94% Mismatches: 263
Query Match: 11.79% Indels: 158
DB: 6 Gaps: 18
US-09-857-669-2 (1-609) x AX033468 (1-1737)
QY 3 IleLysProThrAlaLeuLeuLeuProAlaLeuPhePhePheProHisAlaTyrAlaPro 22
Db 16 CTAAACTCACCACGACATTTTAA---GCATTAAAGTTGCTTCCGCCATTTGCTGAACAA 72
QY 23 AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerPro 42
Db 73 ACCGTTGATATTGAAGTTCGAAGGCATTCGTGTTCCGTGCTGCGTAATACAGAT--- 129
QY 43 AspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSer 62
Db 130 -----CTCAATGTTAAATTTGATTATAA 153
QY 63 GluIleLysAspMetValGluLuhisLeuProleuIleThrGlnGlnGluGluVal 82
Db 154 GAAGAAATGGATGGCTCGAAGCGTATCAACATTTGTAACCAAGCC----- 201
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
Db 202 -----GTGGATCGTGT 213
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAsp----- 119
Db 214 TTGCTGCTGTTTGGTTATTATGAATCTTCGTCGTTTGAACGAAACAGCGTCAAGC 273
QY 120 -----GlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaVal 137
Db 274 AAACGCGATTTATTGATGCTCATGTTACACGAGCGAGCCAAACAAATTTGCGGGACT 333
QY 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTrpArgAsn 157
Db 334 GATGTCAAATTTAGGGGGAAGCCGCAAGATGATAATTTT-----AAT 378
QY 158 AlaLeuGluAsnTrpGlnProValcylSerAspPheAspSerLysAspSerTrpGluAsn 177
Db 379 GCATTGCGCTAAACCTTACCAGAAAGATGGCGTTTGGTTGAGCACCACCACTTACGATGAT 438
QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 439 TACAAACAGCGATTTACAGTTTAGCATTAATCGTGGGTATTTTTCGATGGAACTTTAA 498
```

QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAsp 217
Db 499 ATTTCACGTTTGAATCAGCCCTGAAACCATCAACCATGCGGCGAATGTTATTTGAT 558
QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
Db 559 AGTGGTGTCCGTATTCATATATGGAATATTACTTTTATGCCATTCACAA---ATTCCGGAT 615
QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
Db 616 GATTACTTAATTAATATTCTTAACATCAATCTGGTATCCGTATTATTAATGAATATTG 675
QY 258 LeuAspPheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
Db 676 TCGGATTTTAAACCGAGATTTTCCATCTCTCAAAATGGTTTACGTCAGTATTAGTTCAGCCT 735
QY 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
Db 736 AATGTTAATCAT---AAAGCAAAACACTGGATGTGAGATATTCTTTATCCACGATAA 792
QY 298 ArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle 317
Db 793 AAAATGCGATGGAACCTCGGTGGTGGCTTTTCTACTGATGCGCGCTTCACGACAAATA 852
QY 318 AlaTyr----- 319
Db 853 GCGTGGCAAAACCTTGGATTAAATAGCGTGGACATAGTTTGGTTCAAATCTTTATCTC 912
QY 319 ----- 319
Db 913 TCTGCACCAAAACAAACTCTAGAGCAACTTATCGAATGCCACTGCTTAAATAATCCATTA 972
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLys 339
Db 973 AATTATTACTATGATTTTCCGCTGGTGGAGGGAAGGAAAGAGAAATCATACCAATACG 1032
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
Db 1033 AGAGTGTCTTACGTTGTACGCG-----TTACGTTATTTGG 1065
QY 360 ThrSer-----AsnValSerTyrAsnArgSerThr 369
Db 1066 AATAATGCGCATGTTGGCAATATTTTGGCGGACTTCGTATGCGATACACAGCTTTTACA 1125
QY 370 ThrGlnAsnLeuGluLysArgAlaPhe-----SerGlyGlyIleTyrPyrVal 385
Db 1126 CAAGCGGATATCACTGATAAACCCTTACTCTTTATCCAACTGTTGGATTTACTCGTACT 1185
QY 386 ArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLys 405
Db 1186 CGATTACGTTGGTTCCTTTGCCACTTGGGC-----CATGTCGCAAAA 1230
QY 406 IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrp 425
Db 1231 ATT-----ACTTTGATTATAGC----- 1248
QY 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1249 AAACGA----- 1254
QY 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1255 -----ATTGGCTATCAATCTCTTTTATAAAGTGCAGCACTCT 1296
QY 466 AlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThrPheIleArgGly 484
Db 1297 AGCGGTGGTGGTTCGTACTTATGCAAGAAATCATCGTCTC-----GTTCTCGTGTCT 1347
QY 485 GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 1348 GAAATCGGGTATTTCATACAAAAGGATTGGAATAATCCGCTACACTGCGTTCTTTT 1407
QY 505 SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524

Db 1408 GCTGGTGGCATCGTAGTGTGCGGGTACGGCTATATAAAATTTGCGCTATAAATAGA 1467
QY 525 AsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro 544
Db 1468 AATGGAATTTGCTGGTGGTTCGCGATTGCTTACCACTTCTTTAGTAATATCAATCA 1527
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPhe 564
Db 1528 GTCATATCAAAATTTGGTGGCGGCACTTTTGCAGATAGTGGATTAGTCCGATTAATAC 1587
QY 565 LysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 1588 ACAGCAAAAGAGCTCGTTATGGCACAGGCTTGGTGGTGGTGGCAGTGGT 1647
QY 585 ProPheSerPheAspIleAlaTyrGlyHisSerAsp-----LysLysIleArg 600
Db 1648 GCGATTAAATTTGATATTGCCACACCATTCGTGATTAACATACACCAAAATATTCAA 1707
QY 601 TrpHisIleSerLeuGlyThr 607
Db 1708 TTTTACATCGGACTTGGTACA 1728
RESULT 15
U32752/c
LOCUS U32752 11833 bp DNA linear BCT 29-MAY-1998
DEFINITION Haemophilus influenzae Rd section 67 of 163 of the complete genome.
ACCESSION U32752 L42023
VERSION U32752.1 GI:1573692
KEYWORDS
SOURCE Haemophilus influenzae Rd.
ORGANISM Haemophilus influenzae Rd
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE 1 (bases 1 to 11833)
AUTHORS Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
Merrick, J.M., McKenney, K., Sutton, G.G., FitzHugh, W., Fields, C.A.,
Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodek, A.,
Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,
Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Fraser, C.M.,
Smith, H.O. and Venter, J.C.
TITLE Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
JOURNAL Science 269 (5223), 496-512 (1995)
MEDLINE 95350630
PUBMED 7542800
REFERENCE 2 (bases 1 to 11833)
AUTHORS Tatusov R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S.,
Borodovsky, M., Rudd, K.E. and Koonin, E.V.
TITLE Metabolism and evolution of Haemophilus influenzae deduced from a
whole-genome comparison with Escherichia coli
JOURNAL Curr. Biol. 6 (3), 279-291 (1996)
MEDLINE 96398784
PUBMED 8805245
REFERENCE 3 (bases 1 to 11833)
AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 11833)
AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REMARK The H. influenzae sequence has been updated by R. Fleischmann. New
database matches have been assigned, product names have been
improved, and a number of frame shifts have been corrected. We
gratefully acknowledge the work of Tatusov et. al. We have
incorporated their annotation into the /notes fields of the
corresponding H. influenzae genes


```

Db 10083 GCGATTAAATTTGATATATGCGACACCCATTCGTGATAAGATACAGCAAAAATATTCAA 10024
|||||
QY 601 TrpHisIleSerLeuGlyThr 607
|||||
Db 10023 TTTTACATCGGACTTGGTACA 10003
|||||
RESULT 16
LOCUS AE013668 10831 bp DNA linear BCT 26-JUL-2002
DEFINITION Yersinia pestis KIM section 68 of 415 of the complete genome.
ACCESSION AE013668 AE009952
VERSION AE013668.1 GI:21957368
KEYWORDS
SOURCE
ORGANISM Yersinia pestis KIM.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia
1 (bases 1 to 10831)
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Watson, J. S.,
Blattner, F. R. and Perry, R. D.
Genome Sequence of Yersinia pestis KIM
J Bacteriol. 184 (16), 4601-4611 (2002)
12142430
2 (bases 1 to 10831)
Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,
Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,
Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,
Straley, S. C., McDonough, K. A., Nilles, M. L., Watson, J. S.,
Blattner, F. R. and Perry, R. D.
Direct Submission
Submitted (21-SEP-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1. .10831
/organism="Yersinia pestis KIM"
/strain="KIM"
/db_xref="taxon:187410"
247..990
/gene="cysQ"
/feature="cysQ"
/feature="y0654"
247..990
/gene="cysQ"
/function="phenotype; central intermediary metabolism:
Sulfur metabolism"
/notes="affects pool of
3'-phosphoadenosine-5'-phosphosulfate in pathway of
sulfite synthesis; residues 2 to 247 of 247 are 78.45 pct
identical to residues 1 to 246 of 246 from E. coli K12 :
B4214; residues 2 to 247 of 247 are 79.26 pct identical to
residues 1 to 246 of 246 from GenPept : >gb|AA23224.1|
(AE008906) affects pool of
3'-phosphoadenosine-5'-phosphosulfate in pathway of
sulfite synthesis [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="sulfite synthesis pathway protein"
/protein_id="AA084242.1"
/db_xref="GI:21957369"
/transl_table=11
/translation="NMLEQIGQLAREAGVAINAVYQGDPKPLDIAHKKDDSPVTAADIA
AQHIIKAGLARLPDIPLLSEEDPPAWDVRQHWRLVLDPLDGTKEFLNRNGEFTVN
IALIDQEPVLGVVYVPVTEVMYSAANGAKCEGRMQLQVRDAQPPLVYVSRSH
DAELEDYLSQLEGHPTISVGSLLKCLVAEKGAKLYPRFGPTNWDTAAGHAVAIAG
AQVHDWGKTLSTPREPLNFGFRVSLF"
complement(993..1631)
/gene="y0655"
/feature="y0655"
complement(993..1631)
/notes="residues 26 to 211 of 212 are 59.13 pct identical
to residues 1 to 184 of 184 from E. coli K12 : B4216;

```

```

residues 26 to 212 of 212 are 60.42 pct identical to
residues 1 to 185 of 185 from GenPept : >emb|CAD06884.1|
(AL627283) putative exported protein [Salmonella enterica
subsp. enterica serovar Typhi].]
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084243.1"
/db_xref="GI:21957370"
/transl_table=11
/translation="MKNLPTFRGCKPTSLSTRISLQGHNMINKNLLIISIFTPLLVS
AHNPLDQRVPPVGVGDKGELDYDTKTNFNYKNNSQLPKGRVTOHJAGRTSAKE
MNAPLIDAIKQANLPREYOTTTITDDDAIIGTAMFYRNSIESKREPFMSOFTIDS
KGNVAKAWGVTEKSSAIVLVDPQGNVFKVKGCAITSAEVQOVIDQLHLVQQ"
1892..2104
/gene="y0656"
/feature="y0656"
1892..2104
/gene="y0656"
/notes="residues 1 to 66 of 70 are 77.27 pct identical to
residues 14 to 79 of 81 from E. coli K12 : B4217; residues
1 to 66 of 70 are 75.75 pct identical to residues 23 to 88
of 90 from GenPept : >gb|AA23226.1| (AE008906) putative
cytoplasmic protein [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084244.1"
/db_xref="GI:21957371"
/transl_table=11
/translation="MRIFQYNPARKVAMVYKTLFRGRLYIKMGAFBDMGKILVPKV
KDRHFEVMAEVNRQVLQAQDTLLT"
complement(2212..3543)
/gene="y0657"
/feature="y0657"
complement(2212..3543)
/gene="y0657"
/function="putative transport"
/notes="residues 1 to 437 of 443 are 82.15 pct identical to
residues 1 to 437 of 447 from E. coli K12 : B4218;
residues 1 to 443 of 443 are 81.39 pct identical to
residues 1 to 446 of 447 from GenPept : >gb|AA23227.1|
(AE008906) putative hemolysin-related protein [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AA084245.1"
/db_xref="GI:21957372"
/transl_table=11
/translation="MLNSILLILFLIAVSAPFSLSEISLAASRKIKLKLADGDTNA
LVFLKQETPGMFTYVQIGLNAVATIGVGDAAFSFKLVFERPMAPELDAQSF
VGSFVLVTSFLFLFADLTPKRIGMTSPKAVAVRIVNPMRFLMIFRPLVWFNCGANL
IFRLFLKPLVNRDDITSDDIYAVRAGALAGVLRQEHLLIENFELESRTVPSMSTS
REWYIFDLRESDSDIKSTPHSKFLVCGHIDQVGVYVDSKDLNRLVGNOSLV
LSSGVQIRSAIIVPDTLTLSALESFKTAGEDFAVILNEYALVVGITITLNDVMTLMG
DLVCGOGEQIVARDENSLIEGGPTIEDVMRVLHIDDFPQSGNYETIGGFMVYMLRK
IPKRTDFRYAGYKFEVVDISYKIDQLLVTLRLSDQAPILPKAPHSSDA"
complement(3821..4459)
/gene="msrA"
/feature="msrA"
complement(3821..4459)
/gene="msrA"
/function="enzyme; proteins - translation and
modification"
/notes="residues 1 to 210 of 212 are 73.33 pct identical to
residues 1 to 210 of 212 from E. coli K12 : B4219;
residues 1 to 211 of 212 are 72.51 pct identical to
residues 1 to 211 of 212 from GenPept : >gb|AA23228.1|
(AE008906) peptide methionine sulfoxide reductase
[Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="peptide methionine sulfoxide reductase"
/protein_id="AA084246.1"
/db_xref="GI:21957373"
/transl_table=11
/translation="MONVONTAVIDAANALPGLTSPVSPLIHAVGHSMYTIPEGMD
LAFAMCCFWGAERLFWQPGVYSTAGYSGGHTPNPTTYHEVCSGRTGHAIEVVRVFD

```

PAVTSYQQLQIFEWENHPAQGMROGGDVGTOYRSAIYVLTPEQEEQAHKSRERFQA
MEKAGDRVITSETVALPYAYEDDHOQYLUHKNPHGCGGLGGVGVCLPPNV"

gene
4734..6470
CDS
/gene="y0659"
/note="y0659"
residues 7 to 578 of 578 are 71.20 pct identical to
residues 7 to 577 of 577 from E. coli K12 : B4220;
residues 7 to 578 of 578 are 71.32 pct identical to
residues 7 to 577 of 577 from GenPept : >gb|AAU23229.1|
(AE008906) putative outer membrane protein [Salmonella
typhimurium str.]
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084247.1"
/db_xref="GI:21957374"

/translation="MLRYPILCVCLLATPIAYAAVNRLOVEGLSGDLERNVVARLS
TIGTDEVTAGRSRSDVAIRQGLRALGYDPTITFELONRPAAPASVLIKVVYGE
PVLIAGVDIYVQGAFTDPDQALVRRDTPKGVNLHGDFDNTSSYLGALRRGYF
DANIKSOLGVAQCLHEAFWDIDFSGQRYREGKVIPOGSDIREYIQLNLPFHEGEF
YTSDELAELNRRLAATNFWNSVYVSPFOADAKESKILPDADVTPRTNVTVELGGVA
TDVGPRTASWRKPMNNSCHSLTTTALSAPEQTLDFSYRPLLRNPVEQYLIQGG
FRRTDLNDTNSDTLNVARFWDLSGMQRAINLRWLSLDFHFTQGRVDTTMLKPLPGVS
INRTQRGGAMPVNGDSORYSIDVSTTWGSDVDFGIFQAOVMTIRTLGKNRNVAYQ
NVGTIENNRDVPPLSRFPAGDSIRCYKFRNLSPRDSGKLTGASKATGSLVQ
YNFTGRWGAFFVGDSEAVNNSFKSLKGTGAGVGYVRASPGVPKILDIAPIGNETH
GVQPIYGLGPFL"
6467..10384
/gene="y0661"
6467..10384
/note="y0661"
residues 5 to 1305 of 1305 are 57.45 pct identical
to residues 5 to 1259 of 1259 from E. coli K12 : B4221"

gene
6467..10384
CDS
/gene="y0661"
6467..10384
/note="y0661"
residues 5 to 1305 of 1305 are 57.45 pct identical
to residues 5 to 1259 of 1259 from E. coli K12 : B4221"

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084248.1"
/db_xref="GI:21957375"
/translation="MKWVKRLSTAFLLIILLVGALGGLGTTSLHFLINSAAWVP
GLDIASVGGWRGLTITQYOMPVVVKAGOFYLSLQSLKNSLVCNALTAODID
VVVNTQAMPSTPASTPDMGELSTPYPIPLRLSLANVKTIDDTAISLDEPTGA
HKWKSALMPTKINGLLIAKPKTSAGVDPDAVKPAVETMAAKENVAQLSAAPCRSA
TASGOSTASGRANTASGOSATGGGLAVYPAEPATPLGESLKALPAKPLPALPDR
LPLDQIREISGOQLRLTIDNEVLEISLLQASTQDQRIITDLEIKSPQGLSAQQQ
LAKAGPLSLTQSKOLRWPLTGEPOYQNNLRNLNGOAYDYALSTRSDIGTDLPP
AVFTLEGRVQEFNTRLRALAQHTDLTGVDWQRAISWNSVLTSLGINNAKQWP
EWAKLEKTVITRGSIHGSSWQLQVPELDGNVKNRVTARGSLTGNNAQOWHIPGI
MLALGRNKLDMKGLDNLWLDANPOLDCALPGLGVVKGSLKRLNLTPOLLA
DLTANRLWQDELINRIKVDGVRSDQIOGOLAVRVEQLKQADPVVYLSLLTLDARGSE
KOHRLNLWQSDPVSQLSIEGSDFKQQRWRTLNNTREPTPVGEWPLSRANTLDYQ
NSAORTIGPHCNLNAELCAPRIEAGPSQASVYVLRNFDLAMIKFFLPDPTMNG
VFTGRASVWKEGKLPDVRSLSGNVGVQWVGSPPLAFETLNLNGGLANGQVR
ADMVLVNGQVSGQVADPEGRNLSGNVAISNLAMINPLISLNGGLAGILNA
NLRUAGAKSPLYGLKALDKVIDGSMWPFDTIEGRAMNFDGWTSTLEGLITSOG
QLNLSGDADNPLYSAMHARIAKGNLRIPPTARVDVSPDVFETNLPQFTLNGSV
DIPWARTQVEPTAVGVSSNEVNLDOLOPIAPRSASIPISNLRVIRIGNDVRNA
FGLKARQGLDMMQDQGLNGOINIPSGSFRAYGODLVNKGILLRFGSPDQPLL
NIEAIRPDSDTANGVAGVRVTGMADSPBEVFSPPMSQOEALSYLRGGLGNSCA
DSSWMTSKLGMVAGSGKLVCKICEAGVSNALDITOGVSDSSQVSVGVYTKDLGI
KYGVGIFDSIATUTLRILRMLPRDLIEAVSGIDALDLYQFEF"
7056..7247

gene
8.33e-15 Length: 10831
Pred. No.: 366.50 Matches: 152
Score: 36.58
Percent Similarity: 92
Best Local Similarity: 22.79%
Query Match: 11.55%
DB: 161
Indels: 21
Gaps: 21

US-09-857-669-2 (1-609) x AB013668 (1-10831)

QY 5 ProThrAlaLeuLeuLeuProAlaLeuPhePhePheProHisAlaTyrAlaPro----- 22
DB 4746 CCTATTCTCTGTTTGTGTCTATTCTGCCACACCTATCCGCTATGCCCAATATG 4805
QY 23 -----AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAla 35
DB 4806 CGGCTACAGTTGAGGGCTTACGGGTGATTTACAAAGAACATCGAGGCACGTTGTCC 4865
QY 36 LeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPheProVal 55
DB 4866 ACAAT-----GGCACTGATGAAGTTATACCGACATGGCGTTTTCGTCA 4910
QY 56 ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIle 75
DB 4911 CGGGTCGAT----- 4919
QY 76 ThrGlnGlnGlnGluValLeuAspLysGluInThrGlyPheLeuAlaGluGluAla 95
DB 4919 ----- 4919
QY 96 ProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu 115
DB 4920 ---GAAGCTATTTCGGCAGGGCTCGCATCTACTGCAAGCGGTGCGAAACGGATCCT 4976
QY 116 ThrGluLysAsp-----GlyAlaTyrThrValHisIleThrProGly 129
DB 4977 GAATTACAAAACCGCCCGCACCGTTCAGTTCAGTTAATTCAAAAGTGGTTCCTGGC 5036
QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
DB 5037 GAGCGGTATTGATTTGCCGGTCTCGATATCGTACTGCAAGCGGTGCGAAACGGATCCT 5096
QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp 169
DB 5097 GAT-----TATCAGGCACCTGTCGCGCTGATACCCCGCAAAATTCGCTCGGTC 5144
QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
DB 5145 CTCAACCATGGTGATTTTGATATTTTACCAGTCTCTTACTGTTAGCTTACGCCGT 5204
QY 190 GlyTyrProLeuAlaLysLeuGlyLysThrArgAlaAlaValAsnProAspThrAlaThr 209
DB 5205 GCGTATTCGACGCAATATGATCAAAAGCCAGCTCGGTGTTCCGCTCAATTCACAGG 5264
QY 210 ValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIle 229
DB 5265 GCGTCTCGGATATTGATTTCGACAGTGGGCAACGTTACCGCTTTCGTAAGTTATTTTC 5324
QY 230 ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGly 249
DB 5325 CAAGGCTCAAA---ATCCGAGAGGATTATCTGCAGAAATTTAGTCGCTTCATGAGGG 5381
QY 250 ThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHis 269
DB 5382 GAATTCATATACCTCTGATGAACCTGCGCGAGTTCAACCGCGCTTGGCAGCCCACTGG 5441
QY 270 TyrSerGlyAlaSerValGlnAlaAspPhe---AspArgLeuGlnGlyAspArgValPro 288
DB 5442 TTTAACCGTGGTGTGTTTCCCGCGATTTCCAGATGCCAAAGAGAGTAAATTTTACG 5501
QY 289 ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp 308
DB 5502 CTTGATCGGTCGTTTACCCCA-----GCCACTGAG 5531
QY 309 SerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLysGly 328
DB 5532 AATACCGTCGAACCTGGGGCGGTATGCAACCGAT----- 5567
QY 329 TyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIle 348
DB 5568 ---GTCGCT-----CCGCGCTCTGCCGCCCACTGG 5594

US-09-857-669-2 (1-609) x AB013668 (1-10831)

```
OY 349 SerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 5595 CGTAAAGCCT-----TGGATCACTCTCTCGGCATAGCTTAAAC 5633
OY 369 ThrThrGlnAsnLeu-----GluLysArgAlaPheSerGlyGlyIleTrpTyr 384
Db 5634 ACCACTACCGCGCTTTCGGCAGCTGAGCAGACACTGGATTCTCAGTACCGCTTCCCTTA 5693
OY 385 ValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly--- 403
Db 5694 TTGAGAAC-----CCCTCGAGCAGTATTACCTGATACAAAGGGGCG 5735
OY 404 ---ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet----- 420
Db 5736 TTTAGACGT-----ACCGACTTAAACGACACTAATTCAGATACACGACCTTCAAC 5786
OY 421 -----LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
Db 5787 GTTCACGCTTTTGGGATCTCTCCAGCGCTGCAAGGGCGATT-----AACCTG 5837
OY 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
Db 5838 CGTTGGAGCCTCGCATCTATTACCAAGGCAGAGTCACGGATACACGATGTTGCTGTAT 5897
OY 455 SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr----- 468
Db 5898 CCGGGGTGAGTATTAACTGTACCCCAACGCTGGTGGCGGATGCCGGTCTGGGGGAT 5957
OY 468 -----ACGCAGCGCTATTCTATTGATGTCTGATACGACTTGGGGTCCGATGTTTGGC 6017
Db 5958 AGCCAGCGCTATTCTATTGATGTCTGATACGACTTGGGGTCCGATGTTTGGC 6017
OY 469 PhePheThrProGluAsn-----LysLysLeuGly-----ThrPheIleIle 482
Db 6018 ATCTTCAGCGGCGAAGACGTCGATCGGACCTTAGTGAGTGAAGAAATCGTTTGTGGCG 6077
OY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
Db 6078 CGTGCAATGTGGGTGGTGGATCGGACCAACCAATATTTTGACCGTGTTCACCGCTCGCTCGT 6137
OY 503 PheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 6138 TTCTTCGCGGGGGCGACCCGATCTCCGTGGCTATAATTCGTAATATTTCACCGCGT 6197
OY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
Db 6198 GATAGTCACGGAATAATTAACCGGGGCTTCCAAAGTTGGCAACGGCTCTCTTGAATATCAA 6257
OY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 562
Db 6258 TATTAACCTCACGGTAGATCGTGGGGCGCGTTTCGTCGATTCGGGCGACCGCGTGAAT 6317
OY 563 AsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
Db 6318 AATTTACGAAGAGTGTGATTAATAACAGGGCGAGGTGTCCGGGTACGTTGGCATCACCC 6377
OY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys-----Ile 599
Db 6378 GTTGGACCGATCAACTGGATTCGGGGCACCAANTAGCGGATACGACACACATCGCGGTG 6437
OY 600 ArgTrpHisIleSerLeuGly 606
Db 6438 CAATTTACATCGGTTTGGGG 6458

RESULT 17
AJ414157/c
LOCUS AJ414157 216050 bp DNA linear BCT 06-JUN-2002
DEFINITION Yersinia pestis strain CO92 complete genome; segment 17/20.
ACCESSION AJ414157 AL590842
VERSION AJ414157.1 GI:15981328
KEYWORDS Yersinia pestis.
SOURCE Yersinia pestis.
ORGANISM Yersinia pestis
```

```
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
1 (bases 1 to 216050)
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,
Karlisyev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.
Genome sequence of Yersinia pestis, the causative agent of plague
Nature 413 (6855), 523-527 (2001)
21470413
11586360
2 (bases 1 to 216050)
Parkhill,J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES
source
location/Qualifiers
1..216050
/organism="Yersinia pestis"
/strain="CO92"
/db_xref="taxon:632"
/note="biovar: Orientalis"
86..940
/gene="YPO3402"
/note="panC"
86..940
/gene="YPO3402"
/EC_number="6.3.2.1"
/note="Similar to Escherichia coli pantoate--beta-alanine
ligase PanC SW:PANC_ECOLI (P31663) (283 aa) fasta scores:
E(): 0, 72.0% id in 282 aa, and to Vibrio cholerae
pantoate--beta-alanine ligase Vc0591 TR:09KUD1
(EMBL:AE004144) (293 aa) fasta scores: E(): 0, 59.0% id in
283 aa"
/codon_start=1
/transl_table=11
/product="pantoate--beta-alanine ligase"
/protein_id="CAC92632.1"
/db_xref="GI:15981329"
/db_xref="SPTREMBL:O82BK7"
/translation="MLIETPLLRQOIRNRQEKRIALVPTMNLHGHMTLVDEA
KTRADVVYVTFVNPQLQFERPDDLHAPRTIQDEKLRHGADLVFPAPAAADYYPAG
LEKQTVYDIPALSTILEGASRPHGVSITVSKFLNIQPDVACFGKQEQDQLALIR
KMYADGYDINTVGYPTVRANDGLASRNGYLTPEEROTAPOLSKITMWALEKMLAG
ERQIDALLEEAAAQLLRVGTPTDELFTIRDAETLQPLTVDSQAAVILMAALGKARLID
NQLVDLRH"
1351..1731
/gene="YPO3403"
/note="panD"
1351..1731
/gene="YPO3403"
/EC_number="4.1.1.11"
/note="Similar to Escherichia coli aspartate
1-decarboxylase precursor PanD SW:PAND_ECOLI (P31664) (126
aa) fasta scores: E(): 0, 78.6% id in 126 aa, and to
Pseudomonas aeruginosa aspartate 1-decarboxylase precursor
Pa4731 TR:AA08117 (EMBL:AE004887) (126 aa) fasta scores:
E(): 1.5e-27, 58.2% id in 122 aa"
/codon_start=1
/transl_table=11
/product="aspartate 1-decarboxylase precursor"
/protein_id="CAC92633.1"
/db_xref="GI:15981330"
/db_xref="SPTREMBL:O82BK6"
/translation="MIRTLQKLRHVKTQADLHYEGSCAIDQDFLEAAGILEYAI
```

```

DIYVNDGORSFYTAIAERGRSIIISVNGAARACACVGDKLIIICSYVQMSYAARLHH
PKVAYEGENQJQRKAKVPQVA"
/misc_feature
/genes="YPO3403"
/notes="Pfam match to entry PF02261 Asp-decarbox, Aspartate
decarboxylase, score 264.30, E-value 1.6e-75"
complement(1880..3073)
/genes="YPO3404"
complement(1180..3073)
/genes="YPO3404"
/notes="Similar to Escherichia coli hypothetical protein
Yade SW:YADE_ECOLI (P31666) (409 aa) fasta scores: E(): 0,
65.4% id in 382 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC92634.1"
/db_xref="GI:15981331"
/db_xref="SPTREMBL:Q8ZBK5"
/transl_table=11
/translation="MTTERDSEVSLGVEHVPVGEIKEGQLIQVTPVIAEYVEKFG
NGTGFIDKDLREINAKQNNMDLGNLPNQNIITMREAKVYLAADVRSEQFAVL
AENLRYPVIGKIDRLSNWTQVNGIDRLGVSSADAEIDNGIPILTYHHMLKNEENK
RFLNTSTTSDVAFSNOMAYLKAGYDTISLYOLEGVLNNKINLPKVVVLTDFDGLK
SVSRVAYPILKNGFRATAFIISRRIKRHPKQKWPDSIQFMSIAELKGIDQFDIOSH
TFLPHEDNRHPILLSSYHNIIPDERSERLAQNPVHIVLSYPRGFGNCTAVNA
AKNAGHLAVTQCKVKGPDNPYTLKRLYLITRDSIPTMAERIANQPGVPVAVKPAPA
VLEMD"
complement(3400..4170)
/genes="YPO3405"
/notes="yadh"
complement(3400..4170)
/genes="YPO3405"
/notes="Similar to Escherichia coli hypothetical 28.5 kDa
protein in hpc-pand Intergenic region Yadh SW:YADH_ECOLI
(P36880) (236 aa) fasta scores: E(): 0, 80.1% id in 256
aa, and to Vibrio cholerae putative permease Vc0590
TR:Q9KUD2 (EMBL:AE004144) (256 aa) fasta scores: E(): 0,
66.0% id in 256 aa"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, integral membrane
protein"
/protein_id="CAC92635.1"
/db_xref="GI:15981332"
/db_xref="SPTREMBL:Q8ZBK4"
/transl_table=11
/translation="MTRYIWLQSIWIKETRFARIWIQTLPVPVITMSLYFVIRGN
LIGARICMGGRDYMQFIVPGLIMAVITNAYSNVAASFYAKFQSFISIELLVAPVPT
HIVIGYGGVGVARIGVIGLVTIISLPFVPLHVSMSIALTLITLAILSLGGLLN
AVFAKTFDDISLVPFVLTPLTLYLGGVPSLSLPPFQWAVSKLNPTVYVMSIGFRYGE
LGITDVSLATYIGVLVVFVAVFYAWAWYLIERTGLRT"
complement(3409..4164)
/genes="YPO3405"
/notes="Pfam match to entry PF01061 ABC2_membrane, ABC-2
type transporter, score -22.30, E-value 5.8e-05"
complement(3427..3492)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3526..3639)
/genes="YPO3405"
/notes="P500890 ABC-2 type transport system integral
membrane proteins signature."
complement(3592..3657)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3679..3744)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3790..3855)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(3919..3984)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(4045..4110)
/genes="YPO3405"
/notes="one of 6 probable transmembrane helices predicted
for YPO3405 by TMHMM2.0"
complement(4167..5093)
/genes="YPO3406"
/notes="yadh"
complement(4167..5093)
/genes="YPO3406"
/notes="Similar to Escherichia coli hypothetical ABC
transporter ATP-binding protein Yadh SW:YADG_ECOLI
(P36879) (308 aa) fasta scores: E(): 0, 81.8% id in 308
aa, and to Vibrio cholerae ABC transporter ATP-binding
protein Vc0589 TR:Q9KUD3 (EMBL:AE004144) (305 aa) fasta
scores: E(): 0, 71.4% id in 304 aa"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, ATP-binding protein"
/protein_id="CAC92636.1"
/db_xref="GI:15981333"
/db_xref="SPTREMBL:Q8ZBK3"
/transl_table=11
/translation="MTYALEITQLTKTYMGVQVQLGIDLHVEAGDFYALLGPNAGACK
STTIGIISLVNKTSGKVQVGYDIDNDIVNAKRLQGLVQEFNFPEFVQLVITQ
AEGYVTRRDALARAEKYLQSLDLSKRDRAIRLSGMRKRLMARLHPEKLLIL
DEPTAGVDIELRSMWGLKLNAGQTTIILTHVLEAEMLCRNIGIIONGELVNT
TMKOLLSKLESETFTFDLGAKSPKPLKGYGYRLTDTSTLEVDVKRQGLNLSFSQLN
VOCVQVQSMRNKANRLLELFVTLVNGHEBKA"
complement(4458..5003)
/genes="YPO3406"
/notes="Pfam match to entry PF00005 ABC_tran, ABC
transporter, score 198.80, E-value 8.5e-56"
complement(4641..4685)
/genes="YPO3406"
/notes="P500211 ABC transporters family signature."
complement(4959..4982)
/genes="YPO3406"
/notes="P500017 ATP/GTP-binding site motif A (P-loop)."
5329..5991
/genes="YPO3407"
5329..5991
/genes="YadP"
5329..5991
/genes="YPO3407"
Alignment Scores:
Pred. No.: 4, 23e-13 Length: 216050
Score: 366.50 Matches: 152
Percent Similarity: 36.58% Conservative: 92
Best Local Similarity: 22.79% Mismatches: 262
Query Match: 11.55% Indels: 161
DB: 1 Gaps: 21
US-09-857-669-2 (1-609) x AJ414157 (1-216050)
QY 5 ProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyrAlaPro----- 22
Db 137007 CCTATTCGTCTTTTGTGTCTTATTCGTCGCCACACCTATCGCCTATCGCCCAATGTG 136948
QY 23 -----AlaAlaAspLeuSerGluAsnlyAlaAlaGlyPheAla 35
Db 136947 CGGTCACAGTTGAGGGTCTTAGCGGTGATTAGAAAGAACGTCAGGCAGCGTTGTGCC 136888
QY 36 LeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPheProVal 55
Db 136887 ACAATY-----GGCAGTGAAGTTACCGCAGATGGCGGTTTCGTGTTCA 136843
QY 56 ArgTlleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIle 75
Db 136842 CGGGTCGAT----- 136834

```

Qy	421	-----LeuThrAlaSerTrpIysArgGlnIleuLeuAsnValLeu	434
Db	135966	GTTCACAGCTTTTGGGATCTCTCCAGCGCGCTGGCAAGAGCGGAT	135916
Qy	435	IIISProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrLeuGlyThrPheLeu	454
Db	135915	CGTTGGAGCCTCGATCAGTTTACCCCAAGCAGATCCAGGATACCAGATGTTGCTGTAT	135856
Qy	455	SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr	468
Db	135855	CCGGGGTGAGTATTAATCGTACCCCGCAACGTGGTGGGGGATGCGGCTCTGGGGGAT	135796
Qy	468	-----	468
Db	135795	AGCCAGCGCTATTCTATTATGATGCTGATACGACTTGGGGGTCGGATGTTGATTTTGGC	135736
Qy	469	PhePheThrProGluAsn-----LysLysLeuGly-----ThrPheIleIle	482
Db	135735	ATCTTTACGGCGCAGACAGCTGGATCCGGACCTAGGTGAGAAAATCGTTTGTGGCG	135676
Qy	483	ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet	502
Db	135675	CGTGGCAATCTGGGTGGGATCGAAACCAATAATTGACCGTGTCCACCGTCCGTCGT	135616
Qy	503	PheArgSerClycAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla	522
Db	135615	TTCTTCGCGGGGGCGACCGCAGTAGTCGCTATAAATTCGTGTAATTTCCACCGCT	135556
Qy	523	GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlyTyrGln	542
Db	135555	GATAGTCACGGAAATAATCAACGGGGCTTCCAAAGTTGGCACGGGCTCTCTGTGAATCAA	135496
Qy	543	LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla	562
Db	135495	TATAACTTCACGGGTACAGTGTGGGGGCGGTTTTCTGCGATTCGGCGAGCGGTGAAT	135436
Qy	563	AsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro	582
Db	135435	AAATTACAGCAAGTAGTCATTAATAAAGCAGGGCAGGTTCGGGGTACGTTGGGCATCACCC	135376
Qy	583	LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys-----Ile	599
Db	135375	GTTGGACCGATCAACTGATATCGCGGCACCAATAGCGGATACGAGACACATGCGCTG	135316
Qy	600	ArgTrpHisIleSerLeuGly	606
Db	135315	CAATTTTACATCGGTTTGGGG	135295
RESULT	18		
LOCUS	AE008906	24892 bp	DNA linear BCT 31-JUL-2002
DEFINITION	Salmonella typhimurium LT2, section 210 of 220 of the complete genome.		
ACCESSION	AE008906.1	GI:16422950	
VERSION	AE008906.1		
KEYWORDS	Salmonella typhimurium LT2.		
SOURCE	Salmonella typhimurium LT2		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella		
REFERENCE	1 (bases 1 to 24892)		
AUTHORS	McCllelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W., Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.		
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2		
JOURNAL	Nature 413 (6958), 852-856 (2001)		
MEDLINE	21534948		
PUBMED	11677609		
REFERENCE	2 (bases 1 to 24892)		
AUTHORS	The Salmonella typhimurium Genome Sequencing Project		

TITLE JOURNAL	Direct Submission Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA COMMENT Supported by NIH grant 5U 01 AI43283	CDS	841..1236 /gene="rpsF" /note="similar to E. coli 30S ribosomal subunit protein S6 (AAC77157.1); Blastp hit to AAC77157.1 (131 aa), 99% identity in aa 1 - 131" /codon_start=1 /transl_table=11 /product="30S ribosomal subunit protein S6" /protein_id="AAL23211.1" /db_xref="GI:16422953" /translation="MRHYELVPMVHPDQSEQVPMIERYSAITGAGCKTHRLDWGR ROLAYPNKLKHAHYVLMVPEAQEVIDELETTFRENDVIRSMVMRTKHAVTEASPM VKADERRRRDDFANETADDAAGDSE"
COMMENT	Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER: http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/ EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/ , and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.Pangeasystems.com/ecocyc/ The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone. Location/Qualifiers 1..24892 /organism="Salmonella typhimurium LT2" /strain="LT2; SGSC 1412; ATCC 700720" /db_xref="ATCC:700720" /db_xref="taxon:99287" /note="LT2" complement(95..383) /gene="yjfY" /note="STM4389" complement(95..370) /gene="yjfY" /note="similar to E. coli orf, hypothetical protein (AAC77156.1); Blastp hit to AAC77156.1 (91 aa), 71% identity in aa 1 - 91" /codon_start=1 /transl_table=11 /product="putative outer membrane protein" /protein_id="AAL23209.1" /db_xref="GI:16422951" /translation="MRARIMFLAALLPGITATAVELNNHQARNMDDVRSGLVIYIN HNFAESEANLALNEADVRNAMYVHVILIREPGSNGNIIHASANIYR" complement(378..383) /gene="yjfY" /note="putative RBS for yjfY; RegulonDB:STMSIH004274" 560..770 /gene="STM4390" 560..565 /gene="STM4390" /note="putative RBS for STM4390; RegulonDB:STMSIH004275" 564..770 /gene="STM4390" /note="hypothetical protein" /codon_start=1 /transl_table=11 /product="putative cytoplasmic protein" /protein_id="AAL23210.1" /db_xref="GI:16422952" /translation="MQKSAITTVSISLPPQAVYSGMYEVNISAACDGTGVTFTGI NNCTITLSLPAGASIIIRHFSADL" 828..1236 /gene="rpsF" /note="STM4391" 828..833 /gene="rpsF" /note="putative RBS for rpsF; RegulonDB:STMSIH004276"	gene RBS CDS	1547..1789 /gene="rpsR" /note="STM4393" 1547..1552 /gene="rpsR" /note="putative RBS for rpsR; RegulonDB:STMSIH004278" 1562..1789 /gene="rpsR" /note="similar to E. coli 30S ribosomal subunit protein S18 (AAC77159.1); Blastp hit to AAC77159.1 (75 aa), 98% identity in aa 1 - 75" /codon_start=1 /transl_table=11 /product="30S ribosomal subunit protein S18" /protein_id="AAL23213.1" /db_xref="GI:16422955" /translation="MARYFRRRKFCRFTAEGVQIEDYKDIATLKNYITGESKIVPSRI TGRFAYQROLRRRAIKRARYLSLLPYTDRIQ" 1819..2280 /gene="rplI" /note="STM4394" 1819..1824 /gene="rplI" /note="putative RBS for rplI; RegulonDB:STMSIH004279" 1831..2280 /gene="rplI" /note="similar to E. coli 50S ribosomal subunit protein L9 (AAC77160.1); Blastp hit to AAC77160.1 (149 aa), 95% identity in aa 1 - 149" /codon_start=1 /transl_table=11 /product="50S ribosomal subunit protein L9" /protein_id="AAL23214.1" /db_xref="GI:16422956" /translation="MQVILLDKVANLGSIGDOVNVKAGYARNFLVPKGVAPATKKNV EYFARRAELEAKLADVLAANAARAETKINALETVTIASKAGDEKILFGSIGTRDIADA VTAGVDVAKSEVRLPNGLVLTGHEVNFQVHSEVFAKVIINVAE" 2417..3354 /gene="yifZ" /note="STM4395" 2417..2422
FEATURES	source		
gene			
CDS			
RBS			
gene			
RBS			
CDS			
gene			
RBS			


```

/genes="yif2"
/notes="putative RBS for yif2; RegulonDB:STW51H004280"
2428. .3354
/genes="yif2"
/notes="similar to E. coli orf, hypothetical protein
(AAC74606.1): Blastp hit to AAC74606.1 (266 aa), 25%
identity in aa 20 - 250"
/codon_start=1
/transl_table=11
/product="putative permease"
/protein_id="AAL23215.1"
/db_xref="GI:16422957"
/transl_table="MDTQASPFARKNNVYVCAAFCCLLWGSAYPAIKSGYDLFQIA
TDIPKIVFAGVRFIFAGGLLLFALLORPKPIGRFRPQPAQLTLGLTSLQVLF
FYGLAFTSGVKGIMNATGTFPSVLLAHFIYONDRLSYNKTIGCIIIGFAGVMVNVY
NGLDFSNLPGESVYVLAFAILSAATLYGKRLSOTVDPMVNTGYQLIGGLVLVIGY
VFGTSLTIHGFSVVALVYLLTLLSVAPALMSILLKYNRVGMIAFPNPLIPVSAALS
ALFEGENILEWYKIALVLVCGSIGIWRVKKVR"
complement(3404. .4052)
/genes="yifB"
/notes="STM4396"
complement(3404. .4039)
/genes="yifB"
/notes="similar to E. coli orf, hypothetical protein
(AAC77163.1): Blastp hit to AAC77163.1 (224 aa), 85%
identity in aa 13 - 224"
/codon_start=1
/transl_table=11
/product="putative cell envelope opacity-associated
protein A"
/protein_id="AAL23216.1"

Alignment Scores:
Pred. No.: 1 08e-13 Length: 24892
Score: 357.00 Matches: 139
Percent Similarity: 39.53% Conservative: 99
Best Local Similarity: 23.09% Mismatches: 262
Query Match: 11.25% Indels: 103
DB: 1 Gaps: 18

US-09-857-669-2 (1-609) x AF008906 (1-24892)

QY 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
Db 16455 CTGAAGTCGAAGGCTATCGGAGAGCTGGAAAAACGTTCCGCGACAGCTCTCTACG 16514
QY 75 IleThrGlnGlnGlnGluValLeuAspLysGlnThrGlnThrGlyPheLeuAlaGluGlu 94
Db 16515 ATT-----CAGAGCGATGAGTGCACGCGGATCGGCGC-----TTTCGCGCCGA 16559
QY 95 AlaProAspAsnValLysThrMetLeuArgSerLysGlyThrPheSerSerLysValSer 114
Db 16560 GTTGATGATGCGATTTCGCGAGGCGCTTAAAGCGTTAGGCTATTACCAACCCACTACAA 16619
QY 115 Leu-Thr-----GluLysAspGlyAlaIleThrVal---HisIleThrPr 128
Db 16620 TTCGACTTCTCCACGCCCTCGCAAGGACGCGAGTATTA-ATCGCCAGGGTTACGCC 16678
QY 128 OGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAs 148
Db 16679 CGGCCAGCGCGTGTGATTCGCGGTACAGAAGTCATTTTACGCGCGCGCGCGCACGGA 16738
QY 148 pGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValClyse 168
Db 16739 C-----AAAGATTATCTG-----GCCTTACTAAAAACGCGTCGCGCAATCGGCAC 16783
QY 168 rAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrAr 188
Db 16784 GTGTGTGAACCGGAGACTATGACAAATTTTAAAAAGTCGTTAACCCAGCGTGTGCGTACG 16843
QY 188 gLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAl 208
Db 16844 TAAAGCTATTTCGACACCGCAATTCAATAAAGTCAGTTTGGGAATTGCTCTGGCGCGCA 16903

```

```

QY 208 aThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGl 228
Db 16904 TCAGCGGTTTTGGGATATTGATTATGATAGCGGTAGCGCTATCGTTCGGACCTGTCCAC 16963
QY 228 uIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnPr 248
Db 16964 CTTCCAGGTTCCGAG---ATTCTGATGAATATTACAAAATCTGTCGCCGTTTAAAGA 17020
QY 248 OGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGl 268
Db 17021 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 17080
QY 268 yHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlnLysArgVal-- 287
Db 17081 ATGGTTTAATCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 17140
QY 288 -ProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLe 307
Db 17141 GCGGTTAAAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 17200
QY 307 uAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnly 327
Db 17201 CTCACCGCAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17242
QY 327 sGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGl 347
Db 17243 ACCGTGGATGAATCTTAC-----GGTCATAGCTGACCCAGCACCCAG 17287
QY 347 yIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnAr 367
Db 17288 TATTTCTCGCGCGGAACAG-----GTCTGGATTGATTATATAAT 17329
QY 367 gSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAs 387
Db 17330 GCCGCTGCTGAAAAAC----- 17345
QY 387 pArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly-----Argly 405
Db 17346 -----CCGCTGGACCACTACTCTGTCACAGCGCGGTTTTAAACG 17386
QY 405 sIleProGlySerAspIleAsp-----LeuGlyAsnSerHisIleThrMe 420
Db 17387 TACCGATTGTAATGATGACGAGCAGGACTCGACACCGCTTCGCCGTCTTACTGGGA 17446
QY 420 tLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHl 440
Db 17447 CCTTCCAGTGTGTCAGCGCGCGGATT-----AATCTGCGCTGGAGTTTCGACCA 17497
QY 440 sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIl 460
Db 17498 CTTTACCCAGGGAACGTCACCAATACCACCATCTGCTCTACCCGCGCGGTGATGATCAG 17557
QY 460 eArgThrSerAlaArgAlaGly-----TyrPhePh 470
Db 17558 CGAACCCGCTCGCGCGGGGAGCTGATGCCAACCTGGGCGGATTCGCGCGCTATTCACT 17617
QY 470 eThrProGluAsnLysLysLeuGlyThr----- 479
Db 17618 AGACTATTGCAATACGGCTCGGGCTCCGACGTGGATTCTTCGCGCGGTGATGATGAT 17677
QY 480 -----PheIleArgGlyGlnAlaGlyTy 488
Db 17678 CGTCTGGATTCCGACTTTGATCGACCGCCACCGTTTGTGATCGCGCAATTTGGCGTG 17737
QY 488 rThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAl 508
Db 17738 GATGAAACCGGTGATTTCGACAAAGTTCCCGCGGATTACGTTCTTCGCCGGGGCGGA 17797
QY 508 aSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVa 528
Db 17798 CCGCAGTATTTCGCGCTATAAATACAAATCTATTTCGCCCTAAAGATAGCGACCGCACT 17857

```



```

/gene="STV4600"
/notes="Similar to Escherichia coli prophage P2 Ogr protein
SW:OGRK_ECOLI () (72 aa) fasta scores: E(): 4.8e-15, 55.7%
id in 70 aa and to Bacteriophage 186 late gene control
protein B SW:VPP_BP186 (P08711) (72 aa) fasta scores: E():
2.1e-17, 64.2% id in 67 aa."
/codon_start=1
/transl_table=11
/product="putative positive regulator of late gene
transcription"
/protein_id="CAD06722.1"
/db_xref="GI:16505374"
/db_xref="SPTREMBL:O821G7"
/translation="MNCKPGCHSHRSFQVTDSTKERYCQCQINCGSTFVTHET
VVRFTVPALVNNPPHTVTSQGHMF"
/complement(2723..3823)
/gene="STV4601"
/complement(2723..3823)
/gene="STV4601"
/notes="Similar to Bacteriophage P2 late gene control
protein D SW:VPP_BP2 (P10312) (387 aa) fasta scores: E():
0, 54.3% id in 370 aa and to Bacteriophage 186 late gene
control protein D SW:VPP_BP186 (P21679) (389 aa) fasta
scores: E(): 0, 54.7% id in 371 aa"
/codon_start=1
/transl_table=11
/product="putative regulator of late gene expression"
/protein_id="CAD06723.1"
/db_xref="GI:16505375"
/db_xref="SPTREMBL:O821G6"
/translation="MNFSSDLFGLNSRPAFSITIEGKDVTTALDARLMSLTLDNRG
FEADQLEDDADGQIVLPRRGATVQALGWKQGPLPKGATVDEIERSGAPDRLT
IRASDPRELTNTRKSHWOTVGVVKEIARHNLKVALCKDLTKDVLHMDQTN
ESDAPSLKARYGAIASVKGNNLFTROCGRTASGKPLPVITIRKAGDSRHTL
ADKAGTGVLSMLHTRFKKRETTQVRHKKVATPREPAKOGDVLVGTDDENVLV
NRTYTAAGNAKAAQWERLQGVASFSLQALAEGRADLYTEMVKVSGFKQPIDDAE
WTITTHTVSPDNGFTTSLELVKIDDELE"
/complement(3820..4305)
/gene="STV4602"
/complement(3820..4305)
/gene="STV4602"
/notes="Similar to Bacteriophage P2 complete genome U
essential tail protein TR:O64315 (EMBL:AF063097) (159 aa)
fasta scores: E(): 2.2e-26, 45.2% id in 157 aa and
Bacteriophage 186 protein F TR:O80317 (EMBL:U32222) (161
aa) fasta scores: E(): 1.1e-29, 49.1% id in 159 aa."
/codon_start=1
/transl_table=11
/product="putative phage tail protein"
/protein_id="CAD06724.1"
/db_xref="GI:16505376"
/db_xref="SPTREMBL:O821G5"
/translation="MMVLGLYFVLMRTVPVQELQYQSRWHAANSRNRPRPQFLG
PDNDPTLQGLVLEVTGQSLALAEQMAELGKAWPLIEGSGTIYGMFVIESLQTK
TEFFESGMSRRIEFTLTLRVDESLDMFGSLSDQLSNLQDSATSAIGNIKNTVGGLL
Q"
/complement(4305..7085)
/gene="STV4603"
/complement(4305..7085)
/notes="Similar to Bacteriophage 186 protein G TR:Q37848
(EMBL:U32222) (812 aa) fasta scores: E(): 1.4e-14, 24.5%
id in 694 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD06725.1"
/db_xref="GI:16505377"
/db_xref="SPTREMBL:O821G4"
/translation="MSDNNLRQLQVVLGAVDKLTFRFNAQAGSKELASAIRQTRDQTK
KLSDAGQLKSLQDLTQSVSRTCALDQARLRQAOMTREMSSLEPPTKKOTQALQW
RAVSRLEKQOQETROMAARAELVRLGLSAGGARETARIARETERYNQLAEQERR
LREVEGRQRKLNAIKAEKTRFLNRLNSLAGNAGAGAVTTCGTLTLLAPYKAYSESEN

```

```

AAQLAGSMRGPGGVAPPEFKINRLVALVALGDKLPCTTADPONTMNTMLRQGHSAQVI
LNGGESAAITLVOLQMTAAAEFAAKLQDATQTSKMDNMLMDVLRQGYAGVDSG
NMLOGFSEISSAMNIIINKKLEAVKTFAPLLVMADQSGMAGESAGNAYRKIFQALDA

Alignment Scores:
Pred. No.: 2,22e-12 Length: 249050
Score: 357.00 Matches: 139
Percent Similarity: 39.53% Conservative: 99
Best Local Similarity: 23.09% Mismatches: 262
Query Match: 11.25% Indels: 103
DB: 1 Gaps: 18

US-09-857-669-2 (1-609) x AL627283 (1-249050)
QY 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
Db 153234 CTGAAGGTCGAAGGCTATCGGAGAGCTCGAAAAACGCTTCGGCACAGCTCTCTACG 153293
QY 75 IleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu 94
Db 153294 APT-----CAGACGATGAGGTGACCGCGATCGGCGC-----TTTCGCGCCGA 153338
QY 95 AlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 153339 GTTGATGATGCGATTCGCGAAGGCTTAAAGCGTTAGGCTATTACGAACCCATATCAA 153398
QY 115 Leu-Thr-----GluLysAspGlyAlaTyrThrVal---HisIleThrPr 128
Db 153399 TTCGACTTTCCTCCACCGCCTCGGAAAGGACGCGAGGTATTA-ATCGCAGGTTACGCC 153457
QY 128 oglyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAs 148
Db 153458 CGCCGACGCGGTGTGATTGGCGGTACAGAAATCATTTTACGGCGGCGCGCGTACGGA 153517
QY 148 pGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySe 168
Db 153518 C-----AAAGATTATCTG-----GCCTTACTGAAACGCGCTCGCGCAATCGGCAC 153562
QY 168 rAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrAr 188
Db 153563 GTGTGTCACACCGGGGACTATGACATTTTAAAGTCGTTAAACCGAGGCTCGCTACG 153622
QY 188 gGlyGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAl 208
Db 153623 CAAGGCTATTTCGACACGGAATTCATTAAAGTCAGTTGGGAATTCGTCGCGCGTCA 153682
QY 208 aThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGl 228
Db 153683 TCAGCGGTTTTGGGATATTGATATGATCGCGTACGCTTCGGCGCTGCAC 153742
QY 228 ulIeThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnPr 248
Db 153743 CTTCAAGGTTTCGCAG---ATTCTGATGAATATTACAAATCTCTTCGCGTTAAAGA 153799
QY 248 oglyThrProTyrAspLeuAspLeuLeuAspPheGlnIleAlaLeuGluGlnAsnGl 268
Db 153800 GGGTATGATGACGAATTCGAAGATTTCGCGGAACATAACCGCGCTTCGCTACCGG 153859
QY 268 yHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVal-- 287
Db 153860 ATGGGTTAATCTGTGTGGTGGCGCGGAATTTGAGAAATTCGCTGCGGTAACAAATCT 153919
QY 288 -ProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgIle 307
Db 153920 GCCGTTAAAGCGGTGTTATCACCGCGNACCGAACACCATGTAACCGGGGTGGGTA 153979
QY 307 uAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys 327
Db 153980 CTCCACCGACGTCGCGCGCGCGTGAAGGCTCGTGG-----AAAAA 154021
QY 327 sGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGl 347
Db 154022 ACCGTGGATGAACCTCTAC-----GGTCATAGCTGCACCAACGACGACCAG 154066

```

QY 347 yileSerGlnProArgAsnTyrArgGlyAsnTyrTrrThrSerAsnValSerTyrAsnAr 367
 Db 154067 TATTTCTCGCGGGAACAG-----GTCTGGGATTTTGTATATAAAAT 154108
 QY 367 gSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrrTyrValArgAs 387
 Db 154109 GCGCTGCTGAAAC-----154124
 QY 387 pArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly-----ArgLy 405
 Db 154125 -----CGCTGGAGCAATACTATCTGTGTACAGCGCGTTTAAACG 154165
 QY 405 sileProGlySerAspIleAsp-----LeuGlyAsnSerHisAlaThrMe 420
 Db 154166 TACCGATTGAATGATACGAGGAGGACTCGACACGCTTGCGGTGCTACTGGGA 154225
 QY 420 tleuThrAlaSerTrrPlysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHI 440
 Db 154226 CCTTTCCAGCGGTGGCAGCGGCGATT-----AACCTGCGTGGAGTTTCGACCA 154276
 QY 440 sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerThrAlaLeuI 460
 Db 154277 CTTTACCGAGGGAACGTCACCAATACCACATCGTCTTCTACCGCGGCGTGATGATCAG 154336
 QY 460 eArgThrSerAlaArgAlaGly-----TyrPhePh 470
 Db 154337 CGGAACCGCTCGCGAGGCGGAGTGTGCCAACCTGGGCGGATTCGACGCGCTACTCGT 154396
 QY 470 eThrProGluAsnLysLysLeuGlyThr-----479
 Db 154397 AGACTATTGCAATACGGCTGGGCTCCGACGTGATTTCTCGTACTCGACGCGCAAAA 154456
 QY 480 -----PheIleIleArgGlyGlnAlaGlyTy 488
 Db 154457 CGTCTGGATTGGGACTTGTACGATCCGACCGCTTTTGTGTGCGCGCCCAATTTGGCGTG 154516
 QY 488 rThrValAlaArgAspAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAl 508
 Db 154517 GATGAACCGCTGATTTTCGACAAAGTTCGCGCGGATTTACGTTTCTTCGCGCGGCGCA 154576
 QY 508 aSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVa 528
 Db 154577 CCGAGTATTCCGCGCTATAATCAATCTATTCCCTTAAGATAGCAGCGCAATCT 154636
 QY 528 lLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgTh 548
 Db 154637 TAAAGCGCCTCAAACTGGCAACGCGATCGCTGAGTACCACTAACGTCACCGGCAA 154696
 QY 548 rLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMeLy 568
 Db 154697 ATGCTGGGGGCGGTGTTTCGATACCGCGGAGCGGTAGTGATTTCGCGGTAGCGA 154756
 QY 568 sLeuLysHisGlySerGlyLeuGlyValArgTrrPheSerProLeuAlaProPheSerPh 588
 Db 154757 TTTCAAAACCGGAGCGGGTGGCGGTGCGCTGCGGTGCGCGGTGGCGCTGTCAAACT 154816
 QY 588 eAspIleAlaTyrGlyHisSerAspLys-----IleArgTrrPheHisSerLe 605
 Db 154817 CGATTTTCCGCTACCGCTCGCGCAAGCAAGCAACGCGTTTACAGTTTATATCGGTCT 154876
 QY 605 uGly 606
 Db 154877 GGG 154880
 RESULT 20
 AX033470
 LOCUS AX033470 1731 bp DNA linear PAT 21-SEP-2000
 DEFINITION Sequence 3 from Patent WO0047737.
 ACCESSION AX033470
 VERSION AX033470.1 GI:10280230
 KEYWORDS
 SOURCE Haemophilus influenzae.

ORGANISM Haemophilus influenzae
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.
 REFERENCE 1 (bases 1 to 1731)
 AUTHORS Ruelle,J.L. and Thonnard,J.
 TITLE Haemophilus influenzae rd outer membrane sequences used as vaccine
 JOURNAL Patent: WO 0047737-A 3 17-AUG-2000;
 RUELE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ; THONNARD
 JOELLE (BE)
 FEATURES Location/Qualifiers
 1..1731
 /organism="Haemophilus influenzae"
 /db_xref="taxon:727"
 BASE COUNT 509 a 322 c 386 g 514 t
 ORIGIN
 Alignment Scores: 4.13e-15 Length: 1731
 Pred. No.: 355.50 Matches: 138
 Score: 38.01% Conservative: 87
 Percent Similarity: 23.31% Mismatches: 238
 Best Local Similarity: 11.20% Indels: 129
 Query Match: 6 Gaps: 15
 DB: 15
 US-09-857-669-2 (1-609) x AX033470 (1-1731)
 QY 79 GlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsn 98
 Db 145 AAAGAAGAAATGACGGGCTCTGAACGCTATCAGCATTTGGTACCAAGCCGCTGATCGT 204
 QY 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLys 118
 Db 205 -----GGTTTCGGTGTGTTGGTTATATCAATCTCCCGTGGCTTTGNACGAAA 255
 QY 119 Asp-----GlyAlaTyrThrValHisIleThrProGlyProArgThrLys 133
 Db 256 CAGCGCTCAAGCAACGCGATTTATGATGCTCTGTTACACCGGCGGCAACAAA 315
 QY 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu 153
 Db 316 ATTGGGGGACTGATGTCAAATTTGAGGGGAGCCGACAGATGAAATTTT-----369
 QY 154 TyrTyrArgAsnAlaLeuGluAsnTrrPrlGlnProValGlySerAspPheAspGlnAsp 173
 Db 370 -----GATGCGCTACGTAAACCTTGCACAAAGAGCGCTTTGTTGTAACACCAA 420
 QY 174 SerTrrPrlGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu 193
 Db 421 ACTTACGATGATTACAAAACGCGATTTTCACGCTTGCATTAATAATCGTGGTATTGAT 480
 QY 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
 Db 481 GGGAACTTTAAATTTACGTTTAGAATCAGCCCTGAAACCCATCAACGATGGTGGCGA 540
 QY 214 ValValAlaSerGlyArgProIleAlaPheGlyAspPheGlyIleThrGlyThrGln 233
 Db 541 ATGTTTATGATAGTGGTGTCCGTATCATTTATGGCAATATTACTTTTAGCCATTACAA 600
 QY 234 ArgTyrProGluIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
 Db 601 ---ATCCGCTGAGCATATCTCAATAATATTCTTAACATCAAACTCTGCCATCCATTTA 657
 QY 254 LeuAspLeuLeuAspPheGlnGlnAlaLeuGluAsnGlyHisTyrSerGlyAla 273
 Db 658 ATGAATAATTTCTCGGATTTAACCCGCGATTTTTCATCTTCAAAATGGTTAGTCACTA 717
 QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProLysValSerVal 293
 Db 718 TTAGTTTACCTCAATGTTAATCAT---AAAGCAAACTGTGATGTGAGATTTATCTT 774
 QY 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
 Db 775 TATCCACGCTAAAAAATCGCATCGTGGCTTTGCTACTGATGCGCGCTT 834

[illegible][illegible]

gene	complement(1492. .3084) /gene="VC2538"		PID:1790680; identified by sequence similarity; putative"
CDS	complement(1492. .3084) /gene="VC2538" /note="similar to GB:D10483 SP:P31549 PID:216487 GB:U00096 PID:1786254; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="thiamine ABC transporter, permease protein, putative" /protein_id="AAF95679.1" /db_xref="GI:9657122" /translation="MNRPTQLGFWAAGVATFVLAALALWQSPFFNLNIWQDAYY RHVTFQSFYQAFWSTVLSGLAIPAHLSORRFGFGRDLLKLFAIMTLQVLVAIFG LIAITGVSGFLAKTLHSVGPLPFLPSIYGLNGILLAHVFNLPYASRLQLQSLAIPSE OHLKCAHLGSRMOKFKWEMPRLRQOLPHVAGLVFLMCTFSFATVMAKGGPRATTI ELATVQATKDFDLOIGALLAIQWMLLCTLLTFLQRFSSHSPMNSAASKYTERWVRD SKRAVNDASHVAVACVLLPLVWVYLAGLASKCLGKVLTCDFHSAVSHSLOVASLA SSLAAGSAMLTSRLRLRAHRRANGVEMATILVTPGLVYVSTGLFLLRNITD VRLAYITVAVNALMALPYVVKTLISQPMWHLAQQLYELLCASLGNMGWRRLVEWRA LRKPLSHAFATGFLLSMGDLISAJALFSGDSQDFRTLPLYLYQLLGSYQMDTAVALAML IMSYGWFYVIERLFAQRPLTKC" complement(3094. .4086) /gene="VC2539" complement(3094. .4086) /gene="VC2539" /note="similar to GB:D10483 SP:P31550 GB:U09984 PID:216488 PID:497658; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="thiamin ABC transporter, periplasmic thiamin-binding protein" /protein_id="AAF95680.1" /db_xref="GI:9657123" /translation="MKFQITAVATLCLVASAASAAPLTIVTYDSFAENGAPAPAE RAFAQCCDQLQFVALDGVSLNRLRLEGSNSKADVGLGDLNLLIEAKQTGLLATH SVDTSKVTLPGNSDNTFVPDYDYGFAYFNKEKLANPDTSLKLEKRDLLKIYQD RTTSPQGLLGMKMSYVQDPAQWKLASRTVTVTKGWSAYSMFLGSGADLVLSY TTSYAILHIAENDAFATANSRSEHYLOVEAAKVKTSKNGELADKFLQFTLSNEFQS VPTGNWYPVTAVELPKGFDTLAVPTQSLSFSAQEAQMRKPWIRENQALSF" complement(3956. .4348) /gene="VC2540" complement(3956. .4348) /gene="VC2540" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF95681.1" /db_xref="GI:9657124" /translation="MTWPAIFRYWLLTRFPITIOPTHGCPPIHPNGAEIKHCLGPAM TAMFAVPEPDNACVGLGELQNGIMDDSLPSSLCRSRRASLEISTDCCGVNPMC CLCSVCTNPHRLRFLRCRMGPRSSD" complement(4348. .4983) /gene="VC2541" complement(4348. .4983) /gene="VC2541" /note="similar to PID:I813468; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="3-octaprenyl-4-hydroxybenzoate carboxy-lyase" /protein_id="AAF95682.1" /db_xref="GI:9657125" /translation="MSMNNRRTITLAWTGASGAPYGLRLLQCLLAADYQVYLLISSAA RVVLATBGLKLPANPEAAQALVEHLGCASDKLVVCKEDWFSVSGSAAPKOMVY CPSAGSAAATAGMSDNLITERAADVLLKRGOLLVLRTPFSTHLNMLKLSHMG VTIPAPAGPYPHQPSIDDLVDPEWVARILDHILGIEQALVPRHGYDQVRGE" complement(4990. .6345) /gene="VC2542" complement(4990. .6345) /gene="VC2542" /note="similar to GB:U14003 SP:P37773 PID:537075 GB:U00096		
gene	6334. .6597 /gene="VC2543"		PID:1790680; identified by sequence similarity; putative"
CDS	6334. .6597 /gene="VC2543" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF95684.1" /db_xref="GI:9657127" /translation="MNVKNAPFLCDWNGSIIPRYAKASAAVORGYNKSKTIKVR SPSVSSIPVNRKSFYNSLQTRCIAQSPISDEDDPIILLNPFVALH" 6651. .7661 /gene="VC2544" 6651. .7661 /gene="VC2544" /note="similar to SP:P09200 GB:X12545 PID:41416 PID:537074 GB:U00096; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="fructose-1,6-bisphosphatase" /protein_id="AAF95685.1" /db_xref="GI:9657128" /translation="MOKMRTLGFIIVEKOHDFPHASGELSSLASIRLAAKIVNREIN KAGLADIIGASGNDNIOGEQKLDLYANEKPKAALARDQVCGVSEEPDEAVAFSK ELNKAKVVLMDPLDGSNDTDMVSVGTIFSTYRVPVPGTPPTQEDFLQPSKQVA ACVYVGSSTMLVYTTGCVNGYTPDPSLGTPLYLSHNRRIRPENKIKYISINEGNYIRP PTGKTIKFCOENPEEGRPTSYIGSLVADFHNLKLGIGIYDTPSTPHSGKLR LLYECNPMALFIEQAGGLSADGARRIMDKPTELHQRFVFPVGSKNMVHKVTELEY PD" 7831. .8361 /gene="VC2545" 7831. .8361 /gene="VC2545" /note="similar to GB:U14003 SP:P17288 PID:450373 PID:537068 GB:U00096; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="inorganic pyrophosphatase"		
gene	7831. .8361 /gene="VC2545"		
CDS	7831. .8361 /gene="VC2545" /note="similar to GB:U14003 SP:P17288 PID:450373 PID:537068 GB:U00096; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="inorganic pyrophosphatase"		
Alignment Scores:			
Pred. No.:	2.13e-13	Length:	14390
Score:	348.00	Matches:	140
Percent Similarity:	40.00%	Conservative:	96
Best Local Similarity:	23.73%	Mismatches:	251
Query Match:	10.97%	Indels:	104
DB:	1	Gaps:	19
US-09-857-669-2 (1-609) x AE004323 (1-14390)			
Qy	65	LysAspMetValGluGluHisLeuProLeulleThrGlnGlnGln-----	79
Db	14153	AAAGATAACGTGGAGCGCTATCTCTCATCGATCGCGGCACAGCACTATTTCACAGTCTG	14094
Qy	80	-----GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAsp	97
		::: :::	

```
Db 14093 CGTTTCAATCTCAACTGGAAGACGATGACC-----GAAGCG----- 14055
QY 98 AsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr---- 116
Db 14054 -----TTGAATGCACCTCGCTACTACCAACCCAGTATCGATTATTACTATGTC 14010
QY 117 ---GluLysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAla 135
Db 14009 TCTGAAGATACCAAGCGCTTTACGCGCGCAGTCACCTTGGCGGAAGTCACCGCTTGGAGC 13950
QY 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 13949 GAGGTGTGATATGTGATCCCGCGTGAAGCGCAAGGCGATCGCGGATTTTCAACGTTTGATC 13890
QY 156 ArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerPheAspGlnAspSerTrp 175
Db 13889 CGCCGCACT-----GCTTTGAGAGTGGATGCGCGCTTAAATCACAGCTGTAT 13842
QY 176 GluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLys 195
Db 13841 GACAACTCAAAATCGGCATTCCTGTAATCTGGCGTTGCAAAAGGCTACTTTAATGGTGAT 13782
QY 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
Db 13781 TTTGAGGCCAGCGCGTGCCTGCGCGCTTTCAACAAAGCGCGCTTATTGGTGTCG 13722
QY 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyr 235
Db 13721 TTTGATACGGGATCCGTTACCTGTTGGAGGACACCGGTCGAAGGCGAGTCAA---ATC 13665
QY 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 13664 GATGAAATTCGCGTGATGTCGTCGCGCTTTCAACAAAGCGCGCTTATTGGTGTCG 13605
QY 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerVal 275
Db 13604 CAAGTCGCGGAGTTTAAACAAAATCTTCCAAATACAGATGTTGTTCTCTCGGTGTTGTT 13545
QY 276 GlnAlaAspPheAspArgLeu---GlnGlyAspArgValProValLysValSerValThr 294
Db 13544 GAGCGGACTTATCCGACTTGGATGAGGGCGGTGAGTTCGCCGATCAAGGTGACCTTAGCA 13485
QY 295 GluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGly 314
Db 13484 CGCAAGCACCAATCAACTGGAACCGGTTTAGGCTATTCACGCGAGTTGGAGTAGCT 13425
QY 315 GlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValVal 334
Db 13424 GGTTCATTAATAAGAAACCTTGGTGAACAGCAAGCGCACAGTTTTTCACAGCAGT 13365
QY 335 TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro----- 351
Db 13364 TTCTCTCTTCTATACGACGACAGACCTTACCGCGGTTTACAGATCCCATCGAAGAT 13305
QY 352 -----ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 13304 GGTATAATGAGTACTATCGCATTCAGTACGGGATGAACCATTTG-----GACAAGCGT 13251
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyIleTyrTrpTyrValArgAspArg 388
Db 13250 GATACCGAGACTTGGAAATCTAATCTATCATCTTGAACGCCATTTGG----- 13206
QY 389 AlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 13206 ----- 13206
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGln 428
Db 13205 ---CAGTTGGAT---GGCGGCTGCATCGCACCGGCTCTTTATT-----CGTTAT 13164
QY 429 LeuLeuAsnAsnVal-----LeuHisProGluAsnGlyHisTyr---LeuAspGly 444
Db 13163 CPTTAGAGAAATATCGTCAGGGGTTTACAGAGATGATAACAGCAATTTTATTGCGCGGG 13104
```

```
QY 445 LysIleGlyThrThrLeuGlyThr----- 452
Db 13103 ATGACTTATACCCGTACGCGACTCGTCCACAGTGGGTTTACTGACTTGGGGCGGATAAA 13044
QY 453 -----PheLeuSerSerThrAlaLeuIleArg 461
Db 13043 CAGACGATTACCTTTGGAATATGCGGATCCAGCATTCGTTTTCAGAACACACGGTATGGGT 12984
QY 462 ThrSerAlaArgAlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThrPhe 480
Db 12983 TTACAAACCGGATCTCTTGTGTACGAACATTTATGCGGCAATCATCCGCGGCC----- 12933
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 12932 TTAGTTCGCGTTGATGTGTGAGGCAACTTGTGTGATGAATTTTGACCAACTCTCGCCCTCG 12873
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 12872 CTGCGTTTCTTTGCGGTGGTGATTAACAATTTGCGCGGATACGGTTCAAGTCTGATTTC 12813
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 12812 CGCAAGATGCGACTGCGCTTTGACAGGGGGCGAAATATACATCGCCACCACTCAATCGAA 12753
QY 541 TyrGlnLeuProPheThrArgThrLeuSerGly-AlaValPheHisAspMetGlyAspAl 560
Db 12752 TATCAATA-CCGTTTAAACCGGCAACTGTTGCGCGGCGATGTTTCATGGATGTCGGTGATGC 12694
QY 560 aAlaAlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPh 580
Db 12693 GTTTAACGAT-----AATCCTGAATGGAAGAAAGGGTCGCTACTGGGATCCGCTGAT 12640
QY 580 eSerProLeuAlaProPheSerPheAspIleAlaTyrGly-----HisSerAspLys 597
Db 12639 TTCGCGCGCTTGGCCCTATTCGCTGCTGATTTTCGCTATGGGCGGTGGATGCGCGCGGCGA 12580
QY 597 slusIleArgTrpHisIleSerLeuGly 606
Db 12579 TGAGTTCAAAATCCACTTCATCTTTAGGG 12552

RESULT 22
AL646064 Ralstonia solanacearum GM1000 chromosome, complete sequence;
LOCUS segment 8/19.
DEFINITION AL646064 AL646052
VERSION AL646064.1 GI:17428340
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
REFERENCE 1 (bases 1 to 201050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choise,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavi,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Siguier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
```


F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES

source

1. 201050
Location/Qualifiers

/organism="Ralstonia solanacearum"

/strain="GM11000"

/db_xref="taxon:305"

206..1447

/gene="Rsc1326"

/note="RS02848"

206..1447

/gene="Rsc1326"

/EC_number="2.6.1.1"

/function="small molecule metabolism; amino acid

biosynthesis"

/notes="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE ASPARTATE AMINOTRANSFERASE PROTEIN"

/protein_id="CAD15028.1"

/db_xref="GI:17428341"

/translation="MKTIQSAKLNVCYDGRGVLEKAKOMEPEGHQIILKNIGNLA
PFQFAPETQDMIRNLPNSGSDSGIFAAKAVMHYTQQQGIKNVTLDDIYLG
GASELALATNALDAGDELLPADPYPLMTAMTSLSGTTPVHYLCDSNGMPDLD
IRAKIIPNTKGIWVNNPTGALYSADALLHDIITAIAREHGLVIFADEVYDKVLPD
RHTMASLSEDIIVTFNSLSKSVRCYGRAGMNVSGDKRPARKDYIEGLNMLSSMRL
CANVPGMAIQTALGYSINDLVAPGCMRRQRDLAVELITAPGVSCVVKPKAALYM
PRLDPSVYIDDDQTFIRQLLEEKVLLVCGTFGNHNSPDHFRVFLPHEDDLREAI
GRTARFLRYRKHGTGIRAA"

1538..2857

/gene="Rsc1327"

/note="RS02849"

1538..2857

/gene="Rsc1327"

/EC_number="1.1.1.3"

/function="small molecule metabolism; amino acid

biosynthesis; threonine"

/notes="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE HOMOSERINE DEHYDROGENASE OXIDOREDUCTASE

PROTEIN"

/protein_id="CAD15029.1"

/db_xref="GI:17428342"

/translation="MKPIKIGLLGTVGGTGLKVLQRNOEIRRRAGRGIEVALIAY
RNVETRAIVQETGASVPVDPVAVPSPEIDIVELIGYDITRELVLRANGKH
VYANKALLAVHGNELFKAAQDKGVMAFEAAVAGGPIIKALREGLTGNRIEWIAGI
INGTTFILSEMRDKLDFGTVLAEARQLGYAEADPTFDEGDDAAHKLTLSAIAFG
IVQFEFATVETIKLEAVDIRYAEELGRIKLLGLTRRAEAGIELRVHPTLVPAIRL
IANVEGAMNAVQGDGVDITLYYCKGAGAEPTASAVADLVDTVRHLSIDAMLOKES
LAFQDPALENTPLLPTEEVSAYLYMRVADETTVLADITRLAESISIDAMLOKES
REGEPTDITMLSHVVYVEKQVNAATAEALPTVLSKVTRLMEELN"

2889..4334

/gene="thrC"

/note="Rsc1328; RS02850"

2889..4334

/gene="thrC"

/EC_number="4.2.99.2"

/function="small molecule metabolism; amino acid

biosynthesis; threonine"

/notes="Product confidence : probable

Gene name confidence : probable

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE THREONINE SYNTHASE PROTEIN"

/protein_id="CAD15030.1"

/db_xref="GI:17428343"

/translation="MOTRSTRGHAEPQSFSLILGGLAPDGGLYLPQYQVQVSAEBLT
RWRDLPYADLAFELRKHFATDIPEDDLRALTRRTYTPVYCNARNGTADITPLHRL
EGEAGTSLCLSGNPTLAFKDMAQQLGNLFYALDRHAELNLIAGTSGDTGSAA
EYAMRGKRGVRFMLSPHRKMSAQFQAQMSLQDNKINFINIAIEGVFDDAQDIKAVSN
DLEVKARQKIGTGVNSINWAVVQVYVFKGYLLAAAPKIGDKVFCFVPSGNGVNCAG
HIAAMGLPIDKLVATNENDVLDFFFTGTVYRVRSAAQTYHTSSMDTISKANFPER
FVFDLLGRQPARLAOLFVDVDEKGGFSMTGPEFDRIAECFVSGRSTHEDRYNTLRD
VFVYGTWDTHTADGVKVARENLYPGVMVYLETALPAKFGDTIRALDRPERPAG
YADIESLPQRFEVMPADADRIRKIYAQTGL"

4361..4957

/gene="Rsc1329"

/note="RS02851"

4361..4957

/gene="Rsc1329"

/function="miscellaneous; hypothetical/global homology"

/notes="Product confidence : hypothetical

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="CONSERVED HYPOTHETICAL PROTEIN"

/protein_id="CAD15031.1"

/db_xref="GI:17428344"

/translation="MAEAFASCLHDIRACRICEGALPHEPRPVVLADRARVLIVQO
APGRIVHETGIPWDRSGDLREWLRIGYDAFYADRRIVAVPMGFCFGCPNGDLP
RACAPQWHRPLLTAMPDLKLLVGGYQARYLDVPRGMPLTEVVRGVPDRAEGRV
LFPLPHSPRNLAWEKHNPWFDAQIVPRLRAALAAAFV"

5037..6296

/gene="moaA2"

/note="Rsc1330; RS02852"

5037..6296

/gene="moaA2"

/function="small molecule metabolism; biosynthesis of

cofactors, carriers; molybdopterin"

/notes="Product confidence : putative

Gene name confidence : putative

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE MOLYBDOPTERIN BIOSYNTHESIS MOEA

PROTEIN"

/protein_id="CAD15032.1"

/db_xref="GI:17428345"

/translation="MSANDAPAVKPSLLTLEALLQLLAAARGLTETETVPTLHANC
VLAQAVSGLVNPPADNTQMDGTAVRAADITAPGTRLVAQRIIPAGQVGRPLQNGENA
RIFTGALIPPGDAVVMQEQCPDPTGTVTVIDHVPEAGEWIRAGGDIYTAGAEILPR
GTRIGLQAQGLAASVAGCAGLQVRRVAVFTGDELAMGPELPGCAIYNKNFTLR
ALLENGCETIDFGIVPDTLQATRTDRLREAAEHDLIITSGVSVGDEHDKPAVEAE
GRNLWQIAIKPKGPLAFGEVRQMGAGPAAPFGLPGNPSVSPFTLILFVRPFTL
RLIGVDVDTPKRIPMRADFALPKGDRRNEFLRARINAGGGLDLFPNQSSGVLSTVWG
DGSLIDNPPNQPIAAGDTVAFLPPAGLV"

6310..6573

/gene="moaB"

/note="Rsc1331; RS02853"

6310..6573

/gene="moaB"

/function="small molecule metabolism; biosynthesis of


```

/gene="SMC04046"
/function="miscellaneous; hypothetical/global. homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC47408.1"
/db_xref="GI:15075854"
/translation="MKTKHGSCHSGRIREFVDIDLEAGTSRNCYSCKLRWGSASV
KPFRLMCEERAGIDGYQGTMSGHRHREFTACGVTPYGHGVEEIGGAFVSINVAQLD
GVDPAEFALPIQYMDGLHNNWNAPEATRH"
4979..5024
/notes="REP (repetitive extragenic palindromic) element;
Predicted by Homology"
/evidence-not_experimental
5027..5084
/notes="Sm-4 OR SMC04584
REPEAT SM-4
Predicted by Homology"
/evidence-not_experimental
5094..5542
/notes="Sm-2 OR SMC04873
REPEAT SM-2
Predicted by Homology"
/evidence-not_experimental
complement(5586..6029)
/gene="azu2 OR SMC04047"
complement(5586..6029)
/gene="azu2 OR SMC04047"
/function="small molecule metabolism; energy transfer;
electron transport"
/notes="Product confidence : probable
Gene name confidence : putative
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE PSEUDOAZURIN (BLUE COPPER PROTEIN)"
/protein_id="CAC47409.1"
/db_xref="GI:15075855"
/db_xref="SPTREMBL:Q92M26"
/translation="MKTKMLIAMLCATSAGQNAEEVVRVEMLNKAADGRVWAFDPDV
IRAQGDVTFVAKDKGHSALMKGAPEGAETWKGKINEIITVTLSPGVNYQCAP
HVGGMIGAIYGEPEANLEAVGINKYPKRSAAAEKIFAETESGG"
complement(6076..6417)
/gene="SMC04048"
complement(6076..6417)
/gene="SMC04048"
/function="small molecule metabolism; energy transfer;
electron transport"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PUTATIVE CYTOCHROME C PROTEIN"
/protein_id="CAC47410.1"
/db_xref="GI:15075856"
/db_xref="SPTREMBL:Q92M25"
/translation="MTHASSRRRIAVALVTCVTIGLSPRMSAEEDKLAGREIFLERSE

```

Alignment Scores:
Pred. No.: 333800
Score: 342.00

```

Percent Similarity: 36.94% Conservative: 102
Best Local Similarity: 21.81% Mismatches: 267
Query Match: 10.78% Indels: 158
DB: 1 Gaps: 24
US-09-857-669-2 (1-609) x SME591792 (1-333800)
QY 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePhePheProHisAlaTyr 20
DB 200124 CTCACATCGTGGCTCGGCCGTCTCGGCCGGTTCCTCGTCGACGACGCCGCGTTC 200183
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
DB 200184 -----AAGATCTTCGGAATCGCTTCTTCTCGAAAGCGC 200216
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIle----- 57
DB 200217 GAAGAAGAAGTCAGGTCATAGAC-----CCGTTTCGTATACGCTGACT 200261
QY 58 -----AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
DB 200262 TTCGAGCGCGGACAGGATGACGAGAGGCTCCGGGAGGCGCTGGAACAGGCTCGCAGCTC 200321
QY 75 IleThrGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGlu 94
DB 200322 GTTCAGGATCAGGAAGAACCGGCTCTCGGGCGGATCTTGGCGCTGGCGATCAAGCGCGCGC 200381
QY 95 AlaProAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer 114
DB 200382 GACCGCGATCGGCTCCTCGCGCTCTACGAAAAGGCGCTTACGGGCGCAGCGTTCAGC 200441
QY 115 Leu-----ThrGluLysAspGlyAlaTyr----- 122
DB 200442 ATCCTGGTCACGGCCAGGACATGTACAGCTACCCCGCGCCGCTTCCCGCAGCGC 200501
QY 123 -----ThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137
DB 200502 CAGCCCGTGGCGGTCGTGTTCTCGGTTGTTCTCCGCGCGC----- 200540
QY 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
DB 200541 ---GCATTCAGCTGGCGCAGGTCAGGCTCGAGGAGAGCGCGCGCGCTC----- 200588
QY 158 AlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
DB 200589 -----GATCCCGCGCGCTACGACCTTAAGCGCGCGCGCGCGCGC 200630
QY 178 SerLys-----ThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
DB 200631 TCGACCTTGATCATCAGCGCGCGGACGATCGTGAACGATCTCAAGAACAGACAGCAGG 200690
QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp 211
DB 200691 CCCCTGGCAAGCTCGCGCGCGCAGCGCTGTTGCGGATCATCGCGCTTCCACGCTCGAC 200750
QY 212 LeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGly 231
DB 200751 GTGAGCATCCGCGCGCGATGGCGCGCGCTCGCGCGCTCGGCAACCTGAGCGTACGCGC 200810
QY 232 ThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrPro 251
DB 200811 GCCAGACGGTCGATCGCGGATTTTCGTCAGGATATTATTCGCGCTGAACACGACGCGCA 200870
QY 252 TyrAspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyrSer 271
DB 200871 TATTCGCGCGGAGATATTCCCAAGCGCGCGCGCGCTCGCGCGTGTGAATGCTTTTCC 200930
QY 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---ArgValProValLys 290
DB 200931 AGCGTCACCATCAAC---GAAGCGCGCGGCTCGCGCGCGCGCGCGATCCCGATGAAC 200987
QY 291 ValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGlu 310
DB 200987 ---GATCCCGCGCGCTACGACCTTAAGCGCGCGCGCGCGCGC----- 200987

```

Db 200988 ATCCAGGTATCGAAGGCAAGCACCGCTATTTCGGTTTCGGCGGACAGGTCTCGACCACG 201047
QY 311 TyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIle 330
Db 201048 GACGGCTTGGCTCCAGGGCTATTGGGCGATCGCAATCTCTCGGACGTCGCCGAATCG 201107
QY 331 GlySerValValTyrAspMetAspLysTyr---GluThrThrLeuAlaAlaGlyIleSer 349
Db 201108 CTTCCGATCGAAGGCTCGGTGATCGCTCGCGGAGACCAACCGATCTCGCGGCTCGAT 201167
QY 350 GlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThr 369
Db 201167 ----- 201167
QY 370 ThrGlnAsnLeuGluLysArgAlaPheSerGlyIleProGlySerAspIleAspArgAla 389
Db 201168 -----TATTCGGCGGCATCTCTTCGCGAAACCCGCGCC 201203
QY 389 ----- 389
Db 201204 TTCGGACCGGCTCGACCTTACCGCAAGCGTCAAGCGCGCATCGTCGATCCGATGCC 201263
QY 390 -----GlyIleAspAlaArgLeuGlyAlaGluPhe----- 399
Db 201264 TATAGTCGCAAAACGTTACCGCGCGCGCGCTTCGAGCTTTCGGCCGCAAGAC 201323
QY 400 ---LeuAlaGluGlyArgLysIleProGlySerAspIleAsp-----LeuGlyAsn 415
Db 201324 AGTTCTCGGTGGCGGAGAGTGGGTGGCGGATCTCGACGACGCTTTCGGCTCGAAT 201383
QY 416 SerHisAlaThr---MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
Db 201384 TCCTATATCACCGCGCGCTCCGCTCGAATATGTCGCGACGCGCATGACAAGCTG 201443
QY 435 HisProGluAsnGly-----HisTyrLeuAspGly 444
Db 201444 ACCCGCGGAGGCGTACAGCGCGCTGATCAATGCCAAGCGGAGCTACGAGATCAGGGA 201503
QY 445 LysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAla 464
Db 201504 AAG-----ACCTTTTTCAGCTCC-----TTCGAAGCTTCCGCG 201536
QY 465 ArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
Db 201537 TCCGCTATTACGTTTGGACCGCAAGACGC-----TTCGTGCTGCCCGC 201584
QY 485 Gln-----AlaGlyTyrThrValAlaArgAspAsn---AlaAspValProSerGlyLeu 501
Db 201585 AAGCTCGCGCGCGCGCTGCTCGTGGCGAGCAACTCTCCGATATTCACGACCGCCG 201644
QY 502 MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu 521
Db 201645 CGGTTCTTCTCGCGCGCGCGCGTGGTGGCGGCTATTCCTACCGAGATCAGCCCC 201704
QY 522 AlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr 541
Db 201705 CGCGACGCGGATGACGAACACTGCGCGCGCGCGCTCTATGTCAGCGCTCCCTCGAAGCC 201764
QY 542 GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGly----- 558
Db 201765 CCATCGCGGCTACCGCATACGATACGTCGCGCTGTCGCTTCATGTCGCGCGTACGTTTCG 201824
QY 559 ---AspAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyVal 577
Db 201825 GATGACGACGCGCGCGATTTCTCC-----GACATCGCGCGCGCGCGGATCGGTCTG 201878
QY 578 ArgTyrPheSerProLeuAlaProPheSerPheAspIleAla 591
Db 201879 CGCTATCGCACGCCATTCGGCGCGCATCGGCTCGACTTCGCC 201920

RESULT 24

AE006218/c

LOCUS

10302 bp DNA linear

BCT 08-MAR-2001

DEFINITION Pasteurella multocida PM70 section 185 of 204 of the complete genome.
ACCESSION AE006218 AE004439
VERSION AE006218.1 GI:12722230
KEYWORDS Pasteurella multocida.
SOURCE Pasteurella multocida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Pasteurella.
REFERENCE 1 (bases 1 to 10302)
AUTHORS May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittem,T.S. and Kapur,V.
TITLE Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
PUBMED 11248100
REFERENCE 2 (bases 1 to 10302)
AUTHORS Zhang,Q. and Kapur,V.
TITLE Direct Submission
JOURNAL Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES
source
1. 10302
/organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
complement(70..1830)
/gene="PM1809"
complement(70..1830)
/gene="PM1809"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03893.1"
/db_xref="GI:12722231"
/translation="MKLLNSKASLTITKVPFVIGSVLLSTVSLAOLLELVKQIGIONTO
LNNRVYVGMIDKEADSGSRHKQIVREAIIDKALRAYGYQSEVFEQIFSQPKPKD
QLAHVIGKIPKIVADDFVLAQEAQDFEITAVTKNIPEKSGILNIHETDNYKSHIQ
KIALSGYFDGDFLVSRLVSPSTQAMWRLDYDSGERYRFEVTFENAOIREDYLRN
MINPKQPYLINDLSTLNNVTSSNMFSSVLMQVPLDDEHKIVNDVLLQPKKNSM
EVGIGASDVGPRLOLGWTKPWNNRHSFRTNLVYSAPKOTLEATKMPLLKPLNY
XYESAGLENNKNDSEFASLSAIRYNHEAGHQHSLGLRVYDSFIOANVKKDL
LYFTASVRTRLQGLFTPGDGTQKLTIDIGRTWMLSDVDFELKMQSSSLWRTYLOH
HRIVTRLEGLWHTKNIETIPPALRFAGGDRSIRGYTKKIAPNNNKKLVGGSRL
TGSFEYQYVYVPMWLKATFADTFLAANOPTTKELRYGAGMGVRWASPVGAIKFDIATP
VRDKNSKNIQFYIGLGAEL"
complement(1980..2609)
/gene="narp"
/note="PM1810"
complement(1980..2609)
/gene="narp"
/codon_start=1
/transl_table=11
/product="Narp"
/protein_id="AAK03894.1"
/db_xref="GI:12722232"
/translation="MSEKMKVLLIDHPLMRGRIGKQLMELAPNEFVADAGNSGEGTS
LAYOTPDLLIIDLNMKGLSGLDLTKALFAEGIDARILITVLSOLKDLPTLIDAGAD
GYLKIDPEPALLISQIRIAQGEVTLSDSIKALIEROMNADPIYSLTDREMDVLRLL
ATGLSNKQIAGQLFISETVVKVHNRLLRLKLVNHSRVAATVLIYFEHGR"
complement(2625..5141)
/gene="cyaA"
/note="PM1811"
complement(2625..5141)
/gene="cyaA"
/codon_start=1
/transl_table=11
/product="CyaA"
/protein_id="AAK03895.1"
/db_xref="GI:12722233"
/translation="MNYDLFSAOKVVEYLDKLRIBRALSGSGSGEFQHFVQLITLLHLI
NHPNLGYVADAPVGIADFVISPQKQYLLTTTVPSEANQSLPSPFSYRSTNAILGVY


```
QY 254 LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 1143 ATTAATGATCTTCTACCTTACCAACAATTTACACGTCGTCAAATTCGTTTACGCTGTT 1084
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
Db 1083 TTAATGCACCTGTTTATGATGAA---GAACACAAAATTCGTAATGTCGATGTATTGCTC 1027
QY 294 ThrGluValArgHisLysLeuGluThrGlyArgLeuAspSerGlyTyrGlyLeu 313
Db 1026 CAACACGCAAAAATAAGCATGGAAGTGGGATTGGCTGGCGCATCTGATGCGGACCT 967
QY 314 GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerVal 333
Db 966 CGCTTCGAATAGGTGGACAAAGCCTTGGATTAAACATCGTGCACACATTTCCGTACC 907
QY 334 ValTrpAspMetAspLysTyrGlyThrTrpLeuAlaAlaGly-----Ile 348
Db 906 AATTTATATGTTCTCCGCCAAACAGACCTTAGACGCCACCTACAAAATGCCTTTATTG 847
QY 349 SerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer 368
Db 846 AAAAACCCGCTCAACTATTATTACGACTACTCGCGCGGTTTAGAAAATCAAAATAAAAT 787
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArg 388
Db 786 GACACAGATCCTTTCCTCTCCTTA-----TCTCAATTCGCTATTGGAATCATGAA 733
QY 389 AlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 732 GCAGGTTGGCAACATCTTATAGGCTTAGCTGCTCGCTTATGAC----- 691
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 690 TCTTTTATTCAGCGCAATGTAAACAGAAAACATGCTGTCTCCCACTGCCTCTGTC 631
QY 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 630 AGACGACCCGTTTACAAGCGGCGCTATTCGCCGACTTGGGGGAT-----ACACAAAA 577
QY 446 IleGlyThrThrLeuGlyThr-----PheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 576 CTCACCATCGATTAGGTCGATACATGGTGGCTGCTCATGTGGATTCTCCAAAATGCAG 517
QY 464 AlaArgAlaGlyTyrPhePheThr----ProGluAsnLysLysLeuGlyThrPheIleIle 482
Db 516 GGTCTCAGTTTATGGTGCAGAACTATTTCAGCACCACCATCGTATC-----GTTACC 466
QY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
Db 465 CGACTAGAACTCGCTGTTACATACGAAAAATATCGAACGCATTCACCCCGCTTGCCT 406
QY 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 405 TTTTTCGGGGGGAGATCGCAGTATTCGTTGTTATCGCTATGCGCTATAAAAAATTCGCCAAA 346
QY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlyTyrGln 542
Db 345 AACAAATGCAGCAAAATAGTTGGTGGCTCGCTTATTAACCGGGTCATTGAATATCAG 286
QY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 562
Db 285 TATCAAGTGTATCCTGACTGGTGGTTAGCCCACTTTCGCGATACGGGACTGCGCTAAC 226
QY 563 AsnPheLysArgMetLysLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
Db 225 CAGTTTACAACAAAAAATACGCTATGCGCGCAGGAATGGGCTACGTTGGCGCTCACCC 166
QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAsp-----LysLys 598
Db 165 GTAGGGGCAATTAAGTTTATGATTTGCTCCACCCACCTCAGACATTAAGACACACTAATAAT 106
QY 599 IleArgTrpHisIleSerLeuGly 606
```

```
|||||.....|||||
105 ATCCAATTTTATATTGGACTAGC 82

RESULT 25
AE008175
LOCUS
DEFINITION
  Agrobacterium tumefaciens str. C58 circular chromosome, section 233
  of 254 of the complete sequence.
ACCESSION
  AE008175 AE007869
VERSION
  AE008175.1 GI:15157816
KEYWORDS
  Agrobacterium tumefaciens str. C58 (Cereon).
  Agrobacterium tumefaciens str. C58 (Cereon)
  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae; Rhizobium.
ORGANISM
  Hinkle.G., Slater.S.C. and Goodner.B.
  Complete Genome Sequence of Agrobacterium tumefaciens C58
  (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
  Disease in Plants
  Unpublished
  2 (bases 1 to 10029)
  Hinkle.G., Slater.S.C. and Goodner.B.
  Direct Submission
  Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
  Cambridge, MA 02139, USA
FEATURES
  source
    1..10029
    /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
    /strain="C58"
    /db_xref="taxon:181661"
    119..2044
    /gene="AGR_C_4742"
    119..2044
    /gene="AGR_C_4742"
    119..2044
    /gene="AGR_C_4742"
    /note="conserved hypothetical protein PA2543 (Imported) -
    Pseudomonas aeruginosa (strain PA01)"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_4742p"
    /protein_id="AAK88337.1"
    /db_xref="GI:15157817"
    /translation="MRTRNPNTGIAYPKAGTALAVALTFALYPAFADAPAFKPLG
    MRLGSEPEVEVINPKYAVTLDADADKSLKSSLENSLLADKDKPASGDLGLLI
    KARDORLIALYENARGIVNVTAGNVDDIIPNPFDPHSTPVPMPVITVTPGR
    FTGLNRELDVTGRNLDYGLIAGDGLAIIRAGNKLDLKAERGLAKLTKE
    AVANHATVDITMAAEGGPVAPLGAVTVTGKVDGDFIRYSLNGGEGYSPEKLR
    KAADRLQLGVFSLTIKEAGTLARDCTIPLTIEVSEKGRHYFGVGAQYSTTEGIGLQ
    GYWHRLFGOESLRTEGSVRIEASVEGMDYSAGITFTKPCMNPRTFTKTSIL
    AKTENPDTYRAKTLTGTAFAVELNDTDAAGLEVOMATEDAFKNGYLTTSIPLE
    FVTRDQKLNPTGEPFASLAAPSYEALNCTFFSFEISITGKGLCAEDRLIMACK
    LSGVLVGGSDIADIPTRRFAGGGSVRGYSQEISYPNAGDRTGGRYVVGSE
    ARIKVDTIGLVFPFDAGVVDGVTDFDSIRAGAGIGLRYATPGPULRDVAMPLEK
    YDGGNNGFVAGIQGSF"
    2087..6250
    /gene="AGR_C_4743"
    2087..6250
    /gene="AGR_C_4743"
    /note="(AY010120) putative periplasmic protein"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_4743p"
    /protein_id="AAK88338.1"
    /db_xref="GI:15157818"
    /translation="MMKTLIRLKLWGYAALCCVVLVLAVLVFGFTPMGARIAANOI
    SSVLTPDOTIEISPPGGLTGRLDNVTLDROGPVARNQIADVNSPLSLIAGTF
    HADHVSAGSIDVERPLPAQOTTGKSGGSLPIEIVDNFSPDPSILGOSILGRAPD
    LTAEGNLKAAODDMRLSLTAHRRNVAADVADAFIPNENVVLKKAEMKEPESGLLA
    TLLSLGTPAVADLNGEPLSSWTGTLRGNVAGNAVYNTVTHLLGGDGRTRRIEAG
    GQPDLLPAPFAPOLFPAWETKLDANATLYPGRIEIGSTLETGLLTASGTVDNPG
    QNDLAANLIGTAGVDFRWPFLNGEIQALINGLDLSLKGGAHAHLNTYVLSRLSLP
    QGRDDYKLTAEASADLNITNRSCTIATVLSVAQSVSPDIDRLVPLPTIKAPLSLT
    SSAIGFNGATLESASVGGTIRNSFDLADNRLTSSVQLFALPATLPPALAEKFDTTIAL
```



```

/protein_id="AAL43594.1"
/db_xref="GI:17741112"
/translation="MQNPKVIEMKKAIDHCAGAGGTRNISGTHYHVLLEQELAD
LHKEALIFTSGYVSNWATLGLGQKIPGLTIFSDALNHSAMIEGIRYGRCEVIMK
HNDLEDAKKAADPNAPKIAFESYMDGDIAPREICDLADRYGAMTYLDEVHA
WYMGRCGGTIAERGLMDRLTIIEGLTKAFYVMGGYITGSAVCDIFRSASFIF
TALPSPSLAAGIAISIOHLKASPERARHODVRKRLGLLDARGIPIMDNPSHIVFM
VGDAAKCKWISDILLDNHGVVQPINVPTVRKRLRITPTPLHTDADIEQLVGLH
OLWSHCALARAVA"
CDS
2849..3151
/gene="Atu2614"
/gene="Atu2614"
2849..3151
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43595.1"
/db_xref="GI:17741113"
/translation="MGRUPSRRAHCNVGSEPMKKAIIIFALVGLSLASCTQTEKASIG
AVSGAVIGGAVTCNVRGAAGAAGGAGGALIGNASEPGYCYRDYQGYRTARCR"
3399..5252
/gene="Atu2615"
3399..5252
/gene="Atu2615"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43596.1"
/db_xref="GI:17741114"
/translation="MTFALYPAFARDAFAKLFQMLWNGSEPEVEVNPVKYATVLD
AADAKSLASLLESLADKRPASGDGLLKKARDRDLIAALYENARYGGTIVN
VTAGANVDLPDPFVHSTPVMVITVPGPKFTLGNVRLEGDVTGRNLDYGLIA
GGDAGSLAIRAGNKLIDLKAEGRPLAKLTKEAVANHAATNTVDITMAEGGPVAPL
GAYVTIGTVDGDFIRYSRNLNGEPEYSEPKRAADRLQGLVFSSTIKFAGTLA
RTITPLTTEVSGEGRHRYFGVGAQYSTTEGTLQGYWGNLFQSAESLRIEASVRI
AEASSVGEQWAGITFTKPGMFPRTFTKSLIAKFNPDVTRAKTLNTEGFRASIAEL
NDPTDAAGLEQWADTEDAFCKNFYLTSPLEAFNRDQDLNTEGFRASIAEL
SYEALNCTFESPEGSITGYKLGAEADRLIMAGKLGVLGGSDLDITPTFRFAG
GGSVRGYQISLPSYNAAGDAGTGRSVYVSGVEARKIVTDITGLPFFDGGVSDGV
TPDFSIRAGAGIGURATPTFPGLVDVAMPLEKIDGNNFNGIYAGIGGSF"
5814..9458
/gene="Atu2616"
5814..9458
/gene="Atu2616"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL43597.1"
/db_xref="GI:17741115"
/translation="MRLSLTAHRNVADEYNADVAFLPNENYVLKAKMKEPEGGL
ATLLSLPCTPAVIDLNGCEPLSSWGTLCRGNVAGNAVNVNTRHLLGDGPTBRIA
GGQOPILLPPAPROLFAGETKLDANATLYPQRIETGSGTLETGALLITAGSYDPN
GQNDLAANLIGTAGVDFRHWPLNGEIQALINGDLKGGADAHNITVLSURSLAL
PQRLDDVLTAESADNTINHSQITATVLSAQSSFSVSPIDRLVRAPLTIKAPLSL
TSSAIGFNATLESASVGTGSLGSPDLADNRLTSVQLFALPATLPALAEKFDITIA
LQGDIDLTGGRTSINLVNIGKSTTEAGNYSLENDALSKLPALKFKPLTPQAKG
IADFATEASGALAAVDFNVITSSSELAILAGRKLEALKNASGRADPAAPKAFPSAGS
LDROKIDANVNVNTENGTAPELHVAVGRNLIKCKLQFSQOFLPTGNLSFNPDLAL
LALAAQAOQDIDAGDIALSNANGKIATIKANGNIRQGTITISKLAAIDISDILKA
LAINGISADSNVACTMAISGLNATIGHSCTTAQFVNGRYDNAPLYVYKSATGSP
MTVRYVTFSAKGPVPLRXPSTIAIQNGTARADLTITIGGRVFNVTAGTSGDID
KADIRSLKANGRAGNLAAGLDAGSISGTYSKAGASDPDSVNLNWNANEAQVTRAAG
LAALQKANGRAGNLAAGLDAGSISGTYSKAGASDPDSVNLNWNANEAQVTRAAG
AAQTAQGLDVGDTAAIDVNTSASAPVVTGTSITDGTRLTDVRLNLTINGLATIK
FDRDRAVIRSLTGKVGGGTISGTSGVAGGSGFPADISITLDRAGYNDGTLVTVV
SGTLLTKGPLLSNAPLQNLTLDRSAITIPKLPASLTIDVYKHKNAPKVRQAQAL
GADQGGSGSSSTINLDLQVNPAGSIFVGRGIDAEALTGNLTIRGTAAPVVISGPFEM

```

```

RGELEILTRRLDFTTGNITFGGGLVPVLDKMDKADSTVGSSTVTVTVSYSGNANDPTFAFSS
APALPDQVMAQIIFQGSMSKLSALQIARLADAAQAGGRSTSLPDKRLNGLGVDL
DISTDSEGGARVSAGKYLNERITYLELOQSGESGAKIINLDVGRGVKLRGEAGNGEG
AAGIFYEKEY"
9612..10103
/notes="chew"
/notes="Atu2617"
9612..10103
/gene="chew"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="chemotaxis protein"
/protein_id="AAL43598.1"
/db_xref="GI:17741116"
/translation="MGRILMAMINSTNFGSTLELIAPRLHDOBFQVKTITIREISGW
APSTPIPHAPKDVIGVNNLRVSVIPIIDLAKHLGKMSKSTVANERSAIYVAVVHMVIGM
LVDRVSILITIPANQVQVPEISAFSKISSEGIANEHGMICFLNLAUKMFKRTEAD
LAA"
BASE COUNT 2178 a 3136 c 2967 g 2020 t
ORIGIN
Alignment Scores:
Pred. No.: 6,1e-12 Length: 10301
Score: 323.50 Matches: 148
Percent Similarity: 36.95% Conservative: 94
Best Local Similarity: 22.60% Mismatches: 290
Query Match: 10.20% Indels: 123
DB: 1 Gaps: 21
US-09-857-669-2 (1-609) x AE009208 (1-10301)
QY 6 ThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyrAlaProAlaAlaasp 25
DB 3381 ACCCGCTTGGCTTGTGTTTACGCTTGCCTGTATCCGGCGTGT-----GCCGCGTAT 3434
QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
DB 3435 GCTTTTCGCTTCAAACTGTCGCATCGCTCTGGGTTTCGGAAGAACGGAGGTCCAA 3494
QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGlu----- 63
DB 3495 GTTATCAAC---CCGGTCAATATGCGCTTGCACGCGCGCGCGACGCCACAAGTCC 3551
QY 64 IleLysAspMetValGluHisLeuProLeuIleThrGlnGlnGlnGluValLeu 83
DB 3552 CTGAAGAGCAGCTCTGGAAGAACAGCTCGCTTCTTCTGGCGCAGCAAG----- 3596
QY 84 AspLysGluGlnThrGlyPheLeu-----AlaGluGluAlaProAspAsn 98
DB 3597 GACAAACCTGCCTCCGGCGATCTCGGCTTCTGATAAAGCGCGAGACGACAGGACAGG 3656
QY 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
DB 3657 CTGATCGCGGCACCTTTATGAAACCGCCGTTACGGCGCATCTGCAACGTGACCGTGGCC 3716
QY 117 -----GluLysAspGlyAla 121
DB 3717 GGAAAGAACGTTGATGACCTGCCTCCCAACCCGTTTTTCGATCATCAGCAGCCTGTACCC 3776
QY 122 TyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIle 141
DB 3777 ATGGTCATACGGTAACGGCGGGCGGCGAAATTCACCTCGGCAAC-----GTTCCGCTT 3830
QY 142 LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161
DB 3831 GAAGCGCATGTA---ACGGCGCGCAATCTTTGATGAATACGGCTGTATCGCC----- 3878
QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
DB 3879 -----GGCGGTGATGTCGGGTCTCTGGCAATTATCCGCCGCCGCCCAACAG 3923

```

QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAla 201
 Db 3924 CTGATCGATGATCTGAAGCGGAGCGCGCGCTCCGAAACTACGAAACGGCGGCG 3983
 QY 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
 Db 3984 GTCCGCAACACCGCGCAATACGGTCGACATTTACAACTGCCAGCTGCGCCCGC 4043
 QY 222 IleAlaPheGlyAspPheGluIleThrGlyGlnArgTyrProGluGlnIleValSer 241
 Db 4044 GCGCGCTCGCGCAGTCACCGTCACCGCGGAAACACCGTCGATGGCGACTTCATTCG 4103
 QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
 Db 4104 CGTTATTCGCGCTTAATTCGCGGAGCGCTTATTCGCGGAAAGCTCGCAAGCGCGC 4163
 QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
 Db 4164 GACCGCTCGCGCAGTCGCGCTTTTCAGCGCTCAGCTCAAGAACCGGGAACGGTT 4223
 QY 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
 Db 4224 GCGCGCAGCGACCATTCGCTCACCATCGAAGTCTCGCAAGCAAGCAGCGTTATTTT 4283
 QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
 Db 4284 GCGTCGCGCGCAATATTCAGCAGCGAAGTATCGCGCTTCAAGGATTTGGGTCAC 4343
 QY 322 TyrAsnLeuPheAsnLysGly-----TyrIleGlySerVal----- 333
 Db 4344 CGTAACCTCTTCGACAGCGGGAATCACTACGCGATCGAAGCGTCTCTCGCTATTCG 4403
 QY 334 -----ValTyrAspMetAsp----- 338
 Db 4404 GAGGATCGAGCTCGAAGCATGATATTTCGCGCGGATTTACCTTACCAAGCCGGG 4463
 QY 339 -----LysTyrGluThrThrLeuAlaGlyIleSerGlnProArg 352
 Db 4464 ATGTTCAACCGCCGACGAGCTTCAAGCAGCGCTGATTCGAAGACCGCAACCGGAC 4523
 QY 353 AsnTyrArgGlyAsnTyrTyrTyrSerAsnValSerTyrAsnArgSerThrThrGlnAsn 372
 Db 4524 ACCTACAGGCGCAACCGTCAGCGGACG----- 4553
 QY 373 LeuGluLysArgAlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAsp 392
 Db 4554 -----GCGGTTTCGCTACCAACTGAACGAT----- 4580
 QY 393 AlaArgLeuGlyAlaGluPheLeuAlaGluIleArgLysIleProGlySerAspIleAsp 412
 Db 4581 -----ACGGATACAGCGCGCGGCTCGAGGTCCATGGCGCGATACCGAA 4628
 QY 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
 Db 4629 GATGATTCGCGCAACCAAGATATCTACGACCTCATACCGCTGGAATTTTCGCGCAC 4688
 QY 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThr 452
 Db 4689 ACG-----CGGACGATTAAGCTTAATCCGACCAAGGTTT 4724
 QY 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro 472
 Db 4725 CGCGCATCTTGGCGCGCAAGCAAGCATATGAAGCGCTCAATGGCAGCTTCTTCCTCT 4784
 QY 473 GluAsn-----LysLysLeuGlyThr-----PheIleIleArg 483
 Db 4785 TTCGAAGCTCTATACCGGCTATTAAGGACTTGGTCGCAAGACGGTTGATCATCGCA 4844
 QY 484 GlyGln-----AlaGlyTyrThrVal---AlaArgAspAsnAlaAspValProSerGly 500
 Db 4845 GGAAGGCTCTCCGCGCGTGTTCGTGGCGCGGACGCTACAGGATATACCGACCAACC 4904
 QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520

Db 4905 CCGCGCTTTTTCGCGCGGTGGCGGATCGGTGCGGGCTATAGCTACCAAGAGATTTCG 4964
 QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
 Db 4965 CCGTATAATTCGCGCGGCGGATCGGACTTGGCGCGCTCTAGCTGTGGTTCCTTCCG 5024
 QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetCly----- 558
 Db 5025 GCGCGCATCAGGTCACCGATACCGCTTGTCCCTTTTCGACCGCGCGCTCGTT 5084
 QY 559 -----AspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGly 576
 Db 5085 TCTGATGGGTAAACACCGGATTTCTCC-----GATATCCGGCGCGCTGCGGTATCGC 5138
 QY 577 ValArgTyrPheSerProLeuAlaProPheSerPheAspIleAla 591
 Db 5139 CTGCGTTACGCGAGCGCTTCGCGCCTCTCGTCTCGATGTCGCC 5183
 RESULT 27
 Locus AE003957
 DEFINITION Xylella fastidiosa 9a5c, section 103 of 229 of the complete genome.
 ACCESSION AE003957
 VERSION AE003957.1 GI:9106207
 KEYWORDS
 SOURCE Xylella fastidiosa 9a5c.
 ORGANISM Xylella fastidiosa 9a5c
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.
 REFERENCE 1 (bases 1 to 10074)
 AUTHORS Simpson, A.J., Reinach, P.C., Arruda, P., Abreu, F.A., Acencio, M.,
 Alvaranga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S.,
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.,
 Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrer, H.,
 Colauto, N.B., Colombo, C., Costa, F., Costa, M.C., Costa-Neto, C.,
 Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H.,
 Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A.,
 Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R.,
 Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A.,
 Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P.,
 and Marino, C.L.
 TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis
 Nature 406 (6792), 151-157 (2000)
 JOURNAL 20365717
 MEDLINE 10910347
 PUBMED
 REFERENCE 2 (bases 1 to 10074)
 AUTHORS Simpson, A.J.G., Reinach, P.C., Arruda, P., Abreu, F.A., Acencio, M.,
 Alvaranga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S.,
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.,
 Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
 Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
 Costa-Neto, C., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
 Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J.S.,
 Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C.,
 Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
 Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L.,
 Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, P.,
 Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
 Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L.,
 Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
 Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, F.G.,
 Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G.,
 Nunes, L.R., Oliveira, M.A. de Oliveira, M.C., de Oliveira, R.C.,
 Palmitieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira
 Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V.,
 de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V.,
 Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
 Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de
 Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M.,

Tsubako.M.H., Vallada.H., Van Sluys.M.A., Verjovski-Almeida.S.,
Vettore.A.L., Zago.M.A., Zatz.M., Meidanis.J. and Setubal.J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES

source

Location/Qualifiers

1. .10074

/organism="Xylella fastidiosa 9a5c"

/db_xref="taxon:160492"

/clone="9a5c"

complement(271..759)

/gene="Xf1230"

complement(271..759)

/gene="Xf1230"

/note="similar to SPIP76270 (percent identity: 49 %/query
alignment coverage: 87.7 %/subject alignment coverage:
77.6 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106208"

/translation="MPMLAPFLLTGNKPEYQAQEAQAKALHGETDRIRANANLAAL
IFHSLGFWGFTFYDGRGLVGFPGFLACIRIPLDGVCGTGAARTQRIDDDV
TFPDHACDSASRSELVPLLLKDDTLIGVLDLDSPTCARFSDSDGTGIESIARVFSV
LR"

1052..2905

/gene="Xf1231"

1052..2905

/gene="Xf1231"

/note="similar to SPIP39320 (percent identity: 27 %/query
alignment coverage: 95.9 %/subject alignment coverage:
102.6 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106209"

/translation="MKSIRCEWNVREITLSMLPRYTAQMLRRFIPMLSLWLLFASLA
QANGVDKIQIKGLGDGDEEMRNITVLSYVKGEGEPRLEYLLISQAEFFTRQ
ALEPGYAITRIDAPRONDHITVIVYDKGPEVRVQAVMTAAQADHYLQRLD
EDFKPLGIFNHYEASKYRITRLAERGYPDADTHRRVEVTRAAHAADIDITWE
SGRRDMGYFHEHYDFHEGLNPLVWDEGSYFHEGLKLRSLTKLDYSSSIDIQ
PKPEADPCNVPDVKLERAKSIYTAGISYSGSAGLRAGVRRYNNARGHKMA
RLDYAQLKSLPTAYOIPAFKWDGWTFTARAYDEQTKVIDLRNVKLSAARSGOINR
HLTATSLNLRGRNRYRADGCTNTVAYQOSTLVYPOLEASYVDVDDATEPBGSAAT
VILRGASALSKSNFTQHLGQLRWFHGLGASSRLILRGAGTTWASDLVAMPPLSRF
FAGVNSIRGAFREVGPRTKAPDAFALGANHVFSAGAEFYIRKGGPFGGAVFVDSG
SAPNRYPDWHTGVGIGLYRSPGVRVDIARGLNPSDKIQLYIDIGANL"

3184..6723

/gene="Xf1232"

3184..6723

/gene="Xf1232"

/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using
Glimmer/RBSfinder/Start codon shift: 411"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106210"

/translation="WDPOIMRLLLRLRLDVLMSNASLSIGOSEPFNFPKWPVPLP
SINPLALDITITIDNFRISKLOPLIHLIRGAVNVADGELKTRQAVDSRGNIT
RADGYTRQDYTDLTATVLPVSRGPPASFGLVIRGDLRLQLALGRAPAPIQA
DVLNGRSAPRWTFSAASEHPEGLLPGLTTPSATALFNNQAGGGRADLKGQ
FKQCEQIVLOPSRLQANQMLVEPLVIDAFGGTNRGLKRVLDLPDPAIRFHSYVNA
RGLSEFSPSPGNTQVRAVPVVEAGLVSGSTLRAWAAVGRARLTRDAQARLIPDV
RGDAEHATKQVOAATPGGLLDFTGGVTWAPALRWDLKAEALQDFPGYIFPGWNGFVS
GQLTSQGRORTVSSGIEAIVAVRGLNGELRORPMQASRETFODHOGEGAROLIGDS
HISAKGKVPDHLDAVLOQLSHLNDLPSKEGTLKGSLLLRGPAHPDVIIDLNGTSL

RNDYRAKQAFALHGLHPWRGNNGKLTQAOELEIGTLLDSVRIHSGAVENILQADY
RIAMALLTHGVHRDGAWRGDSQRIEIPAKGPAPALRAPAOFVYMAPVETLSATC
LDAQGGGALCLHADMPHEGLVLKGOALPLTLVHPMLPKNGRSINLRGALTLDGHLRP
HGDWEGFRASSLEGIRGEPRYADVLGVQSGELVRYDQFSQVDLTSHIYVKL
GICFOGNGFVAKFTGMDPYAPLTGDLNNMSRLYWFLEFVDICNGRLIEGHVSL
RGTSGPLLGGEATRDFTAEYPSMGLNSGOGREFTAQDPSARILASVSGDGTLLN
IDGGLSWFGENNPLQLTNRNLAYNTSELRTVSDPDLRFLGINNTMOLQCEVAVSE
ADIDLERLDGVSVEDVVLDPDVPTAVSPLENNLRITVLGDKVNMSSGFLKGGLS
QOMQIRARPGHEMTANGLDVGRYKAYGODLILIRQLVWSNNIVSDRPNIRARR
IADITAGIDVSGRAQSPRVQVSDPMSQSEALSTIVGIRKLTGTSASDENQQVASAA
ALSAGGLIASQLGAKLGLDDAGISQSSITLGGSVLFGKLYLSPKLYVGVGVSMVSGGS
VLTLYLIRSGFDLESSTVETKGSVNRKEK"

gene

CDS

complement(6826..6945)

/gene="Xf1233"

complement(6826..6945)

/gene="Xf1233"

/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106211"

/translation="MFLFEHEHIGNSAMSLAIGASFLVLDQPGFIGQAMRCRP"

7500..8396

/gene="Xf1234"

7500..8396

/gene="Xf1234"

/note="similar to SPIP77541 (percent identity: 62 %/query
alignment coverage: 98.3 %/subject alignment coverage:
99.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="carboxyphosphonoenolpyruvate phosphonmutase"

/protein_id="AAF8404.1"

/db_xref="GI:9106212"

/translation="MSTSLPIQTQCQCFRTALMTERPLQVIGANNAYHALLAQRAGYR
ALYLSGGVAGSLGLPDLGNTMLDVLIDVRITDVCDDLPLLDITDGTGFSALNIE
RTIKTLIKAGAAACHIEDQAAKRGHPPGKEIVLSLDMADRVKAAADARTDAAPFLI
RFTDALQAEVESAEIQAQVVEAGDSIFAFAVSULETYHFVEVIRVPLANMTEF
GTPPLTRQLSAGVVALFPLSNFRANKAETVYTAIRRGHQQAVLPHMQTRA
LYERLDYHAFPEQRDLALSARST"

complement(8703..8778)

/gene="trNA-Lys-1"

complement(8703..8778)

/gene="trNA-Lys-1"

/product="trNA-Lys"

/note="found by tRNAscan"

9291..9446

/gene="Xf1235"

9291..9446

/gene="Xf1235"

/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106213"

/translation="MLPLAGSGTTGVAALRAGHRFIGMEISLGSYSDVAKORLPDRVCG
FIRDVYL"

9570..9887

/gene="Xf1236"

9570..9887

/gene="Xf1236"

/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF8404.1"

/db_xref="GI:9106214"

Qy	82	ValLeuAspLysGluGlnThrGlyPhe-----LeuAlaGluGlnAlaProAsp	97
Db	1286	GTGGTTGTAAGAGCAGGGCGAATTCGGCGTGGAAATATTATTCGTACGACCAAGCCGACATTC	1345
Qy	98	AsnValLysThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeu-----	115
Db	1346	CAGACCCGCAAGCGCTGGAGCCCTTGGCTATTATGCTCCACCATTCGCATTGCATGCC	1405
Qy	116	----ThrGluLysAspGly---AlaTyrThrValHisIleThrProGlyProArgThrLys	133
Db	1406	CGCGCCCAAGATGATCATCATCACCGTGGTTATTACGTGGCAAGAGAGAGCGCGTCCGG	1465
Qy	134	IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGlu	153
Db	1466	CTACGTCAGGCGCAGCTTGGCATGACAGGAGCCCTCGCGAGAGC-----CAT	1513
Qy	154	TyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAsp	173
Db	1514	TATCTTCAGCGTGACCTGGGAAGATTTCAAGGCTAAGCTTGGCGAAATATTCAACCATGCC	1573
Qy	174	SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu	193
Db	1574	ACGTATGAAGCCACGAAGTTCGCATACC CGCGCTTGGCTGAGCGAGGCTATTTCGAT	1633
Qy	194	AlaLysLeuGlyAsnThrArgAlaLalValAsnProAspThrAlaThrValAspLeuAsn	213
Db	1634	GCGGATTTTCACCCATCTCGCGGTGCGAGGTGACCGCTGCCCGCATCGACGACAGCATTCAT	1693
Qy	214	ValValAspSerGlyArg-----ProIleAlaPheGly---AspPhe	227
Db	1694	TTGATCTCGGAAGCGGCGCCCTATGACATGGTCCGCTTCGCTTCCTACGACTAT	1753
Qy	228	GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln	247
Db	1754	-----TTCCACGAAGGGCTATTATTCACCTGGTGCTACTGGGAC	1792
Qy	248	ProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn	267
Db	1793	GAAGTGACTATTTCACGAGCAGCGCTGGATCGCTCGCTGAGTCCCTGACCAAAATG	1852
Qy	268	GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---Arg	286
Db	1853	GACTATTTCACGACCATTGATATCCCAACCCAAAGCCTCGAGGAAGCCGCTGAAGGCAAT	1912
Qy	287	ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg	306
Db	1913	GTTCCTGGTGTATCATCGCTCGRAGCTGCCAAAGAAATCTACACGCGGCGCATCAST	1972
Qy	307	LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsn	326
Db	1973	TATGCGACGCAAGTGGTCCGCT-----	1996
Qy	327	LysGlyTyrlleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAla	346
Db	1997	-----CTGCGTGCT	2005
Qy	347	GlyIleSerGlnPro---ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyr	365
Db	2006	GGTGTGTGACGCGCGCTACATGAATGGCGTGGACAAAAATGAATGCACGATTGACTAT	2065

AUTHORS	DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.-J., Haselkorn, R., Kyrpides, N. and Overbeek, R.
TITLE	The genome sequence of the facultative intracellular pathogen <i>Brucella melitensis</i>
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
PUBMED	11756688
REFERENCE	2 (bases 1 to 10029)
AUTHORS	DelVecchio, V.G., Redkar, R.J., Patra, G. and Mujer, C.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA
REFERENCE	3 (bases 1 to 10029)
AUTHORS	Elzer, P.H. and Hagius, S.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
REFERENCE	4 (bases 1 to 10029)
AUTHORS	Kapatral, V., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Haselkorn, R., Kyrpides, N. and Overbeek, R.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA
REFERENCE	5 (bases 1 to 10029)
AUTHORS	Letesson, J.-J.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire, Laboratoire d'Immunologie et de Microbiologie, Universite of Namur, 61 rue de Bruxelles, Namur 5000, Belgium
REFERENCE	6 (bases 1 to 10029)
AUTHORS	O'Callaghan, D.
TITLE	Direct Submission
JOURNAL	Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Nimes 30900, France
FEATURES	source 1. 10029 /organism="Brucella melitensis" /strain="16M" /db_xref="taxon:29459" /chromosome="I" /complement(71..1942) /gene="BMEI1895" /complement(71..1942) /gene="BMEI1895" /codon_start=1 /transl_table=11 /product="OUTER MEMBRANE PROTEIN" /protein_id="AAL53076.1" /db_xref="GI:17983939"
gene	translation="MAFEIFGHLWGKDKQPDIDIPKTSYSDVTTTTGDRKNADGKE ADLKSVTEAGSIVSDADKPSAGSAGLLAKARGDRIIRLSALYGRGRYGGTISIKVQD REANDIPDPEIPENNAKVAITVDGPGQFLFSRTAISNTAPPNGRRDKVQVPEAGFA PGQESKGTILKAERLAVAEWQEGYAKARVTEGVDVADHADNRVSDADIALDPCRKAY YGVSVVGTALNQEERKPRRGEGAEVSTIDGCGVTSYWMHRLNLRGERLFAKVSQ EPDGSPLITLNQEERKPRRGEGAEVSTIDGCGVTSYWMHRLNLRGERLFAKVSQ IGSQSDSPDKNTILLGASAKFGVTPVTDVATDQAKREVLDATFETSIWAKTG FTQIFSDLSGALYANASCHFPVDDVFGKRDFTTAGLGNLSDNRNKKPPSSGFYL VGNFQPEYHGNFATRTAGRTVHGFGQDVRVLAGRLKVGSIVGSGADLPDPSQ LFTLAGGSGYGVGRIYVAGNGNIIGGRSLVEANGEVRIKTRIDTSIGAVAFPDAGY VGEKSPDFSEQMRVGGGLRYLTSGLPIRLDVAVPLNRRSGDPNYGYVIGQAF" complement(2475..2990) /gene="BMEI1896" /complement(2475..2990) /gene="BMEI1896" /codon_start=1 /transl_table=11 /product="Hypothetical Membrane Spanning Protein" /protein_id="AAL53077.1"
gene	translation="MNSAGPCRWNCRIGNSGSGWISARYVRFYASVAGGYAAQPPATY IYNNGYDDWGPALWGIGIGWASSNMGYPWGWGGRGPNYIRGCIGRNCSQSG PCRRGWNPHWRGPPORRAQWRGPHVNRPFMGPGFAMGPHFGGFGGGRGCMH IGNMRSLHAGR" complement(3594..3594) /gene="BMEI1897" /complement(3594..3594) /gene="BMEI1897" /codon_start=1 /transl_table=11 /product="Hypothetical Protein" /protein_id="AAL53078.1" /db_xref="GI:17983941"
CDS	translation="MTLNFYKLNKVLCELPATRNKKAPKYPGRHAAFGGREGDKA AMPGRQASYMONPSCGNKAGALPSKVSGLAVETC" complement(3625..4071) /gene="BMEI1898" /complement(3625..4071) /gene="BMEI1898" /codon_start=1 /transl_table=11 /product="CYTOCHROME O UBIQUINOL OXIDASE OPERON PROTEIN CYOD" /protein_id="AAL53079.1" /db_xref="GI:17983942"
gene	translation="MPQVLAPRHHLDRRPHSCLSVGSPVNSAHOHTDHAHAGSLG SYLIGFVLAIVTPVPFALAMGYFTATTAVIATAVQIILVHLIYFLYLDPKSEG GWNMLAVFTGIIILAVILGSIWIMYHLDTNMMPMTDPNVRNP" complement(3990..4619) /gene="BMEI1899" /complement(3990..4619) /gene="BMEI1899" /codon_number="1.10.3.." /codon_start=1 /transl_table=11 /product="CYTOCHROME O UBIQUINOL OXIDASE SUBUNIT III" /protein_id="AAL53080.1" /db_xref="GI:17983943"
CDS	translation="MSAIOHTAAPYLGNDHPAHEDHHDAGSTTLVGVFIYLSMDCIL FSSLFATVAVLAHQFAGGPTGRELFDLQFVLVETMLLLVSLTHGLATLSMYKNRNS VLGRLGVTELLGAAFLVMEIYEPHLLVEGAGDPSRSLSNFTLTGTHGLHITSLGI WILVLSAMLLRDGLTERNTRVWMLSLFWHFLDIINIGVFTLYLLGLV" complement(4623..6602) /gene="BMEI1900" /complement(4623..6602) /gene="BMEI1900" /EC_number="1.10.3.." /codon_start=1 /transl_table=11 /product="CYTOCHROME O UBIQUINOL OXIDASE SUBUNIT I" /protein_id="AAL53081.1" /db_xref="GI:17983944"
gene	translation="MFGKLTLEAIPYHEPIIMVTAAVAGSALAIALAITYYRKWVPL WMDTISVDHRRIGVYIYLAIVMLFRGSDAVNMQAOLASGANGFLPPHHYDLSL FTHAGVIMFFVAVPVLGVMNFVPLQIGARDVAFVPLNSGLFWLTVACAILNLSL GVFEAKTQWLAIPPLSGKAFSPQVMDYIYLAIVMLISGMDTLLSAVNLITIKMRAP GMGMOLPFWYWTALCSNVLVYAAPIILATVLSLLTDLYDFHFFHNDAGGPMHYV NLITINGHPVYIIVLPCGFIYSELIATFSGKRLFGYTSVMTATVATVLSFLVWAHH FTTMGGGASVNSFFGVATMIISITPGAKVFWLITNMGKRVMTPMWITIGFCTFY IGMTGVLVAIVPPADVFLPMLFVLIHAFHNVIISGVLFSGFVGVYVWMPKAFGKLKE RIGKNAFVCLWVFPMLAPMPLYVLGIMGMTRNLNPAWHYLIITAAIGVAVILCG IVKQILQIMVDIRDKLDRTNGDTWGSRTLEWSISSPPFPYFNAETPTVRDHWDA DMKQNGYVYRTEKFARIHMPKNTATGVLGILSIPFGFAMITWHIWLIAIASALAMFY ATAHGFNDRDYVVSADVEQVQVEDARTQLTAQEA" complement(6682..7716) /gene="BMEI1901" /complement(6682..7716) /gene="BMEI1901" /EC_number="1.10.3.." /codon_start=1 /transl_table=11 /product="CYTOCHROME O UBIQUINOL OXIDASE SUBUNIT II"
gene	complement(6682..7716) /gene="BMEI1901" /complement(6682..7716) /gene="BMEI1901" /EC_number="1.10.3.." /codon_start=1 /transl_table=11 /product="CYTOCHROME O UBIQUINOL OXIDASE SUBUNIT II"
CDS	


```

/protein_id="AAL53082.1"
/db_xref="GI:17983945"
/translation="MRMRHQMGHTPRITVSEFAVLALSFAFLAGCQNOVLDPKQGVQ
LEEDLFLFATGMLLVVVPVILATLFFFAKRYATNEKADYQPOWHHSKTELAVMLI
PMATVVLGVTVTWTSTHKLDPYRLESNVKPINVEVALDMKWLFIYPDQGIATVNM
AGPVPVNTKLTSSNLMSFFLPOLGSOVYTPSPNOTKHLHLINROGDPAIGIYQYNS
GQGFQAORFKAHAFDOAGFDKWDVADYKAAAGGDLDRARYASLOVPSKHAQYFNS
YGDKALFHDLVNRCWDCKSVCLDDEMHROQIREARSLRKSPPIDFSQWANDLICTV
OPPLRSQLEN"
gene
8234..9199
/gene="BMEI1902"
CDS
8234..9199
/gene="BMEI1902"
/codon_start=1
/transl_table=11
/product="Molybdopter in biosynthesis enzyme"
/protein_id="AAL53083.1"
/db_xref="GI:17983946"
/translation="MIRGELPLDRADETILAHITINDALSURKGTIVLARSIDIAALROA
GIDRVLAARLESIGDIEDAAALAIAGLLAGNIEPGPATIGRYNLFASPDGVFCTDTA
GIDAINHDPRESIVATIRNHARVEAQVATIKIIFPAVPGVLIBETIRAVVHGRPLQ
VEKFRNQGLIOSRLPSIRETVLDRTELHMKRVTRNGGAVTQERVAHDEALATA
IAKVGSAQDLIVIFSASVADEADI VPRAIRRAGGEILIRIGMPDVPDQVLLVGRKGRK
YIIAPCSARSARENSLDRLRLMAGMELDATHLRMGVGGLLL"
complement(9244..9852)
/gene="BMEI1903"
CDS
complement(9244..9852)
/gene="BMEI1903"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C-552"
/protein_id="AAL53084.1"
/db_xref="GI:17983947"
/translation="MNSTNRKFIIMAFATVVFVMTTILSDSLFHSAPKPKGFAIE
VAQSGGGEAAAPKAEKESVSIATLQSDAPARGVEVFRQCQCHTGDGKANKYGVPHL
WGVVRDPIAAYVEGFSVSAAMKFEAGNKNDFELHKLFTSPKGFIKGTAMGFADKK
DNERADLIATYLTSDNPVPLPAADAAADSDKKPAEAAPAN"
BASE COUNT 2368 a 2856 c 2881 g 1924 t
ORIGIN

Alignment Scores:
Pred. No.: 1,12e-10 Length: 10029
Score: 304.50 Matches: 154
Percent Similarity: 37.86% Conservative: 111
Best Local Similarity: 22.00% Mismatches: 279
Query Match: 9.60% Indels: 157
DB: 1 Gaps: 29

US-09-857-669-2 (1-609) x AEO09623 (1-10029)
QY 7 AlaLeuLeuProAlaLeuPhePheProHisAlaTyAlaProAlaAlaAspLeu 26
Db 1993 GCATGCTGCTGCGCGCGCATTTTCCTCCTCGCGCTGCGCT----- 1946
QY 27 SerGluAsnLysAlaAlaGlyPheAlaLeuPhe-----LysAsnLys 40
Db 1945 -----GCATTGCGGTTTGAGATTTTGGCATTTCCACTATGGGCAAGGACAA 1898
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIle----- 57
Db 1897 AGCAAGACCCGACATCATGATCCCAAG---ACCTATAGCTGATGTCACCCACT 1841
QY 58 -----AspThrGlnAspSerGluIleLysAspMetValGluGluHis 71
Db 1840 GCGGATCGCAAGATCGCGACGGAAGGAGCGACCTGAATTCAGTTATCGAAGCGCA 1781
QY 72 LeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeu 91
Db 1780 TCGGGTCTGGTTTCGGATGCGGACAGCCCGCT-----TCCGGCTCGGGGCTGCTG 1727
QY 92 AlaGluGluAlaProAsp-----AsnValLysThrMetLeuArgSerLysGlyTyrPhe 109
Db 1726 GCCAAGCGCGCGGCGATTATCGCGGTATCCTGTCGCGCTTTTATGTGTAAGCGCGTAC 1667

```

```

QY 110 SerSerLysValSerLeu-----Thr 116
Db 1666 GCGCGCACCATCTCCATCAAGTCGATGGCGCGGAGCGCAACGATATTCGCGCGCACG 1607
QY 117 Glu-----LysAspGlyAlaTyrThrValHisIleThrProGlyPro----- 130
Db 1606 GAATTTCCCAATAATGCGANAGTTCGAATACAGGTTGATCCGGGGCGGCANATTCCTGTTTC 1547
QY 131 ---ArgThrLysIleAlaAsnVal----- 137
Db 1546 TCGCGCACGGCTATTTCAATATCGCCGCCGCCGCAACAGCGCGCTATAAGTGCAG 1487
QY 138 -----GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Db 1486 ACCCGCGAAGACGCGGCTTTTGCA---CCGGGCGAGCAAGCAAAATCCGCGACGATCTG 1430
QY 153 GluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGln 172
Db 1429 AAGCGGAAACGCTTTCGCGTCAAGCCTTGGCGTCAG----- 1394
QY 173 AspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro 192
Db 1393 -----GAAGCATATGCC 1382
QY 193 LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeu 212
Db 1381 AAGCGCGCTGTCACGGCGAGGATGCTGTCGCCGACCATGCCGATAACCGCGTTTCCGCG 1322
QY 213 AsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 1321 GATATTCGACTCGATCCGGTAGGAAACCTATTATGGCGCGTACGCGTGGTGGCAC 1262
QY 233 GlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyr 252
Db 1261 GCGCGATGACCCCAATTCGTGCTGGATGACGGCGCTGAAGCGCGCTCAGGAATAT 1202
QY 253 AspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyrSer--- 271
Db 1201 GATCCCGAC-----GATATTGAAAATGCCAAGAGCGCTCTGGCGCGCATGGAAGTC 1151
QY 272 ---GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp---ArgValProVal 289
Db 1150 TTCGCGCGCATACCTTCGAGAACCGCGCAAGATCAACCGCATGCGACTCTGCGCAT 1091
QY 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSer 309
Db 1090 ACGCTCAATGTTTCAGGAGCGCAAGCGCGCGCTTGGCTTCGGCGCTGAATATCCACC 1031
QY 310 GluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyr 329
Db 1030 ATCGATGGCTTCGGTGTGACGAGTACTGGATCCACCGCAACCTGTTGGCGCGCGGAG 971
QY 330 -----IleGlySerVal-----ValTrpAsp 336
Db 970 CGCTCGCTTCGATCGAAGGTGACGGCATTTGGCGCTCGCAGGATAATTCCTTCGAT 911
QY 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyLysSerGlnProArgAsnTyrArgGly 356
Db 910 CCCAAGAATATACCTATCTCTGGTGCCTCTTCGCCAAGCCCGCGCTCTATAGCGCG 851
QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 850 GAC-----ACGATTTTGTGCG-CACACTCGATGTCACCGCAACGCGGATGCTGACGC--- 801
QY 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArg-----AlaGlyIleAspAlaArg 394
Db 800 ---CTATACGGAACCGCATCAATGCCAAGACAGAGGCTTTACGACAGATTTTCAGCGCAGCA 744
QY 395 LeuGlyAlaGluPheLeuAlaGluArgLysIlePro-GlySerAspIleAspLeuGlu 414
Db 743 GTTGTGCGCGCGCTCTATGCAATGCAAG-----CCAGGCGCACTTCCTGACCATGT 690

```

```

QY 414 yAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu----- 430
Db 689 GTTTCGCAAGCGGATTTACGACTCGCGGGCTTGAGCGCAATTTGCTCTATGATAGCGG 630
QY 431 -AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThr----- 448
Db 629 CAACAATAAGCCGACCGCAAGTTCGGGTTTCTATCTCGTGGGCAATATCCAGCCCTTCTA 570
QY 449 -----ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAl 466
Db 569 TGAATTCATTATGCGCAATTTCCGCCACCGG----- 539
QY 466 aGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe----- 480
Db 538 -----TTACCGCTGAAGCGCGCACTATCACGGTTTCGGCCAGACGATCGTGT 489
QY 481 -IleIleArgGly-----GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValpr 498
Db 488 GTTTCGCCAGCGCGGTTGAGGTCGCTCCATCGTTCGGGGTTCGATTCAGATTCGCC 429
QY 498 oSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSe 518
Db 428 ACCAGCAGCAGCTTTCTCGCTCGCGCGCGCGCTCGGTTCCGGCTACGGCTATCGCAA 369
QY 518 rIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySe 538
Db 368 TATCGCGGTACGCGCGGAAACGGGAACATCATCGCGCGCGCTCGCTCGTCGAGGCCAA 309
QY 538 rLeuGlyTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetCl 558
Db 308 TGGCGAAGTGGCACACACCATACCGATTCCATCGCGCGCGGTGGCTTCGGATCGCGG 249
QY 558 y-----AspAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerCl 574
Db 248 CTAATGGGTGAAATTCCTTCCCGGATTTCTCCGAG---CAGATCGCGCTCGGGGTGGG 192
QY 574 yLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
Db 191 CGCGCGCTTGCCTATCTTACCGACCTTGGACGATCCGCGCTGATGTCGCGCTCCGCT 132
QY 593 -----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 131 CAACCGCGCTCCGCGATCGAACTATGTTTCTACGTAGGATGAGGACGCGTTT 74

RESULT 29
AF088896/c
LOCUS
DEFINITION
Zymomonas mobilis fosmid clone 42C11, complete sequence.
ACCESSION
AF088896
VERSION
AF088896.1 GI:4511972
KEYWORDS
Zymomonas mobilis.
SOURCE
Zymomonas mobilis
Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
Zymomonas.
REFERENCE
1 (bases 1 to 39013)
AUTHORS
Lee, H.J. and Kang, H.S.
TITLE
Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis 2M4
JOURNAL
Unpublished
2 (bases 1 to 39013)
REFERENCE
Lee, H.J. and Kang, H.S.
TITLE
Direct Submission
JOURNAL
Submitted (31-AUG-1998) Microbiology, Seoul National University,
San 56-1 Shillindong Kwanakgu, Seoul 151-742, Korea
FEATURES
location/Qualifiers
1..39013
/organism="Zymomonas mobilis"
/strain="2M4"
/db_xref="taxon:542"
/clone="fosmid 42C11"
566..1940
/gene="sah"
566..1940
gene
CDS

```

```

/gene="sah"
/codon_start=1
/transl_table=11
/product="succinylarginine dihydrolase"
/protein_id="AAD21533.1"
/db_xref="GI:4511973"
/transl_table="MIVSEVNFDFGLTPHNYAGLSRGNVASALHAGOPSPYPROAALQ
GLEKMHMDMGLTQGVFPLPVRPVTHLHLGLVKGDDKTLKOAARMICGYNFNLLCS
ASSMARNANATVISEFSDHSDGVRHRTANLALMLHRHLEAOTTYAQLNOMFSNCCPFA
MHPFLCCQHFSDGGAANHRITSAHRTGINFIVYGEKNDIYPAKOKLRASQAVARL
GEVPPDLAMFIPORKEAIAKAFHNDVAVANEYVLLAHAEFAFEGDEGWEKTRTAERID
GFPIIDNITLQAVKYSLFNSQITVLKDRMTMALILOPEVKSDPAVWETVNRITISGN
NPKIKAVVDVRESMANGGPACLRRLRVLPSKAALEAVDQRFILTPRWEKLYQLVEN
FWPEKITPDDLVLPELWAKTAVRAHWALTSLWG"
complement(1937..3067)
/notes="hypothetical protein; orf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21540.1"
/db_xref="GI:4511980"
/transl_table="MIPLFLOLMJMDIFDQSKFFLDKMTGKAGKGTLTAKGYASFDA
SDYDALNITLTKATLARADEMKSTVSGNDITSNRKGSGIKGLKLPTETRYRFIMD
NKETVHDLSCVGRKGOVEDTAHOTKESSASMLASAKLDIRISAPDOLPFAVMGL
ESEHEDLVRGTATAPNITGDMRVIRGTYTFAGRRPDIHGEIQFTGGPPNPALDOI
TAAVTDDITATVKGFGFANKPETITSSPSLAQDEILSRILPGSSVTSUSAVPAQOL
AANLNLRTGGKGFDPDLKLSRISVIGIDRLRVVVGANSSTGQGTSLAAGKYLFKNVMYEV
ITDHTGATQIRISLRTLSLLSEASSFGSSSNVSLRYSKDY"
complement(3317..6106)
/notes="orf2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21536.1"
/db_xref="GI:4511976"
/transl_table="MTDQNYNPERDENSSDSIKEKNPNIVAOPENTEANPKTSFPO
LKHILMGCVFLTGIALFIALILFLDSTGHRLLVRYINRYTSEGLKVIQRIDGS
IYQAFRLVDIKAYDQHRFLDAPSLFVEMRPDLRYGIIITKLHRAVIERLWPELK
PTKSNPAPLDLIDHDIQIDHLRIDIKLNDQGNHRLHIGLALIKGQITVKAEA
GADKQDISGGDTLSQIDAKPTANLLVVEAHLHAPVGVVDGLTRLGKPLTMDTACK
GNQDQWGYIOAGCGDKILLDMQIHNGKQGSVHLGDAVYPSLLPDPLOAFKFKGLQ
IDLTHNKNVDIRGATGEMFRLTALGQVDKMSRNNMSVDLAIPNSOKFLGCDL
SFKNFTSSVLSGLVSPNIDWRLKAGQIGWQEKIAENIQSRGMVLDDGDLHSPMLG
SVKNGTDLPLVNLLENLKEIGAINYGDHLVRDHSISSRYTHANTITLIDLVTKK
WRAAVRGKIDPYKIDIGRLSLTIDSHLPHEKNGGLGMVGMKINGDQKPAKAP
GGEAVLRTDMIANAMGNIRPRLFSLASPEHLEGGNGYQPKGNILFVRGKSKYGAM
TLDVGGIVSHPTIKATAGLTVNGNLAOTKSNLYHQGFELNGRLDGIIRUTEQATO
AVNADIAIDRAHFVYGQPIDISGRILRGKAVPSTDPAFQVEALLNIRYNDFLLKQAA
GRAFYQKEGCGGLTQGNLPAGASPPFNVAIHSSPSADVIRANAGGNVAHLNWHM
AAPVEVRSGRNYDISKVIQPKHIDLSGHYGCNSYAAALQIADIDLSANLLWPE
INLSQLGSVANADLGNIPQISARLALNNFTRVSVGLSQPVLNVVAANTHDKNLGVS
EAVIRQKQTMIGQIR"
complement(6106..8322)
/notes="hypothetical protein; orf3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD21534.1"
/db_xref="GI:4511974"
/transl_table="MLFRHYFRQIHRICLLSVGLISLSAHAEDVSTPKTDSGNKKAAS
PNAIDPFDPFPHLPTLSQPSIEASKDCSSHLPLRTSESSQOGENSGRNVKALDT
ETVSDPNKLLANSADSEBETQAQIETTSVTDLPALSPDLPSHPNMDHMEVLAKE
ESDRLRYKVDVKGLSNIDTQAFKNYVLLSNKKNKADSLSVIGARATNDQELINRIUR
SQDGYDKCAALSITPIEKQYNVRYDYNAGSYIKGQNLGTGSEPEERLIRARVNLH
RGPDIATHLEQAENILITLPHYGYPEAKIGDITLDDDETHGDTYLPYNACNVS
YCSLIIVNNKHIVLDKAKHISHIRAFKEGORYDSQWDDLRLAALATSLFSHVSEPIA
TGRKDEGSEIVLDVRQGRKKAHSIAVTGGTGTGEGFAKQGSWISRIYFPPEGSLTP
SGILTRQOQSALFNRNNAGARDVIOIGLTAGRERYDAYNGYSFGSLSRQSTQ
LMOKRWTSYISGAELTQNTNERSYDFSRERLTRYLTAALPGQGLDYDRSNLNPRTKY
RLNRYFSPETSIGVGLRGYVRLFDASGYYPVADNVNVLAGRFYVASIEGASVQLAPS
RRIYAGGGSVRGYQOGLPKDPYKDAVGLSEELAFEVRYRFGNGLVFPFIDAGQ"
VVEDAIPTFNLREGTIGLARGVYATGPPRIDLATPIARQGESRI SVYISIGQAF"
complement(8322..9562)
gene

```

```

CDS
/ gene="cpxA"
/ complement(8322..9662)
/ gene="cpxA"
/ codon_start=1
/ transl_table=11
/ product="Chemotaxis protein"
/ protein_id="AAD21535.1"
/ db_xref="GI:4511975"
/ translation="MLELLLSLVFSEASQTYVADATQAITERVLLVARRLISODRP
SRPMIANFNSREIRFMSRDOIKHINSOSHLYDLQGLSHDSQILPMDIHID
LAPLGRDHLGCIKSDGSLWVQTLVYPNPMPWLNRLTALITAILVIGAGLVGLLI
QAVTPTKLSLAKVAKGDKGHIFETGPKELRHVAQFQNMQRIRHILRRETEML
AAVSHDLTPRLARQLRSTFEBDDIREAVDLOQMATMIKSVLOFLRHGQDQPEVR
LTDVAIETLVNAAKDAGDISYEGSHYDVTTRPLMKRAISNLINAHLYGKKAI
VRFHDENGVYIQVDDAGPGIPEDQEQVLPQFRLDARQNTAGLGLGATDRTV
KRECKLSLTNLQSGGLRAIFLPAQNKIRKDKGKGYVFSARKGLELLRRPK"
/ complement(9737..10462)
/ gene="arcA"
/ complement(9737..10462)
/ gene="arcA"
/ codon_start=1
/ transl_table=11
/ product="aerobic respiration control protein"
/ protein_id="AAD21537.1"
/ db_xref="GI:4511977"
/ translation="MASQTIWVEDDASRLIARVLKENGYPICPVSGSGEMWEELQ
SRTVDLILMVLDPCKNGKLTRELRSNQYNAIPVILSRADDETDRVLEGLGGDD
YIPFPGKELVAVRVAQLRSQAKVADGNFQSHLHFEGWVLDLRRHLSKLSQT
VELSGAEPDLLSLSPLOHAQRVIGRERLLELSRSRSEASDRSIDLVLSRLRKLSSQT
EAGDMIRITIRGVGYMFTVPVSQD"
/ complement(10491..10865)
/ gene="hisp"
/ complement(10491..10865)
/ gene="hisp"
/ codon_start=1
/ transl_table=11
/ product="phosphoribosyl c-AMP cyclohydrolase"
/ protein_id="AAD21538.1"
/ db_xref="GI:4511978"
/ translation="MQNDQDYGNKLNPAWDEKGLITAVLTDAKTGLLLVAMHNKESF
ELSMATKEATFWSRSRKKIWRKGEESHVMKIQEIRIDCDQDTLWKLKVPNGKSTG
AQSCFYRLVEDGHVLTIPKQI"
/ complement(10920..12113)
/ gene="ygcA"
/ complement(10920..12113)
/ gene="ygcA"
/ codon_start=1
/ transl_table=11
/ product="RNA methyltransferase"
/ protein_id="AAD21539.1"
/ db_xref="GI:4511979"
/ translation="MKLIIRLASKGQVTEQGFVPNSVPDDYISDDCKLSEGAHHIE
PVCRRFSCVGCRIQYDAEYKYNFLTDRIAEAFHQAQSLAPVLAHLSPPYRRRV
ALRPAKAKLTKLYNKTSSHQVDIVECPLLDINFLRAAMDRLRFLQKWLAPRSLAQ
IEMTLADQGDICLLVMPFPTLEATITAFAEKGFARLSDQGYGVETWESERV
VTLGAAPVPTLPAHAFLOATKDGEQTLVHMVKEAVDAHFVADLFSGLGTFALESFQK
RVYAAEGMRDAVALAKAAALGKAVEVHRLOKQDELVRFCGILDPFRAGA
KEQTANLAILPNGRIVVYVSCNPATFARDAKTLLLEAGWILHWKVPVQGFWSLHVEWG
LFTKN"
/ complement(12110..13564)
/ note="hypothetical protein; orf4"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAD21559.1"
/ db_xref="GI:4511999"
/ translation="MTLSKLETLPOHRSIRSSALNRPILRLQAWRAEAAWAKKCS
SYVDLMEKACKVKAASITPSEAGNKPALILCGPGNCGGYVARYLAEWGLIPVSVAY
SEPTDAKKAARSLSFKSVTSLDEAGAPLLIDALFVGMSRSLNALLSLTALSXA
ASLRADIVPSGVDTDKGSLSDIPAFDLTIAIGALKFAHRLPSATICGRVAVGDDI
LTIENPDILTPKRLPAATPYDKYSGRYVAITAGAMAGASLLSAVAARAGAYVA
LLGGETISGPMALVHRSREALADERINALVIGFCLGRDELGVQLRSLMALDTSHPLV
LDADALFLLSREGLDCLRNITQPTIITPHEGSEFKRLFEGISGDKLITRVTKAALSGLV
VVLKADTVIATPDGRAAIISQPAVSWLANRAGTGDVLSGICGAMLARGMDPFDAACTAV
WLHGEARLAGPMIADDLIHLHRSARFEVTVAS"
/ complement(13922..14641)
/ note="hypothetical protein; orf5"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAD21560.1"
/ db_xref="GI:4512000"
/ translation="MISLGGNSSILLIVDHASAFVPEIDILGDKRLVDEHVAVDKV
QTLGQCOALQCNCGVFAAOSRLVLIINDENDFGLMPAISDGYVIRHNGTLTADKX
EARLSRFYRPHOHTDAIQORPRLIISLHSTPCLSKDLNCRYPWQIGLIYNDDR
ATQAIPLLKQALHVLVGENEPYSGRIITNTMNRHAENRIPYLLFEIRQDMLEANTI
GNWCAIKKKVNETLISLKNK"
/ complement(14821..15606)
/ note="hypothetical protein; orf6"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AAD21561.1"
/ db_xref="GI:4512001"
/ translation="MPNGYHALETIFAFKDGDILOAEANDTEDNLTITGPSEGLEA
NKDNLVLRATVLRACQPNKIPAGFSTILDKRLPVAAGIGGSADAAMLMIGQHYQ
IPHELILSLANSLGADVPACVDCLVRGECVGEKLTQICDRSLRBEKPLLYNPRVSCS
TPMIEKHWDGVDRCALVDSGSLGARSGRNDLEPPARKILPIGIEVENLQQQKGVN
FSRWSGSGATCFALFDEIRDRTATYKRLNIDHPFWALSSLLR"
/ complement(15964..17607)
CDS
Alignment Scores:
Pred. No.: 1,346-09 Length: 39013
Score: 300.00 Matches: 168
Percent Similarity: 37.50% Conservative: 99
Best Local Similarity: 23.60% Mismatches: 277
Query Match: 9.45% Indels: 169
DB: 1 Gaps: 33
US-09-857-669-2 (1-609) x AF088896 (1-39013)
QY 1 MetMetIleLeuProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 8048 CTAATGCGACGGGTTAAAGCTTTAGACACGGAACCTGTTTCA-GATCCGAGAACCTC 7990
QY 21 AlaProAlaAlaSerLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 7989 TTGGCTTAATTCCTGACGCATCAGAGAAAACGCGCA-----CAGAAT 7948
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 7947 ATCGAAGACAGCAACTTCCTGCTTGCCTTATCAGAACCCACTGCCTGACCTT 7888
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIlePheGlnGlnGlu 80
Db 7887 TCTCACTTTATGCTGATATG---AATCAACATCTGCTTCCCAAGAAAGATGACGAC 7831
QY 81 -----GluValLeuAspLysGluGlnThr----- 88
Db 7830 CTCGATACAAGGTAGATGTCAAAGGCTCTCAGCAATATCGATCTGATGAGCTTTCAA 7771
QY 89 -----GlyPhe 90
Db 7770 AACTATCTGTTTACTATCGAATAAGAAATAGGCCGATAGCCTATGCGTTATTGTGCT 7711
QY 91 LeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSer 110
Db 7710 AGACCAACCAATGACCAAGAGGTGATTAAACCGGATTTACGATCTCAAGCTATTATGAT 7651
QY 111 SerLysValSerLeuThr-----GluLysAspGlyAlaTyrThrValHisIle 126
Db 7650 GGCAAGCGAGCTCTTTCTTATTACTCCGATAGAAAA---GGGCAGTATAATGCTGCTAT 7594
QY 127 ThrProGlyProArgThrLysIleAlaAsnValGly---ValAlaIleLeuGlyAspIle 145
Db 7593 GAT-----GTCAATGACAGTTCGATCTACAAAGCTAGGACAAATA 7555
```

Qy	453	PhelLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro	472
Db	6582	GGGCTACGTGGGTATGTCGGAATGCTTTTTCATGCG---ACGGGTTACTACCTGTAGCA	6526
Qy	473	GLUAsnLysLysLeu---GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAla	491
Db	6525	GATAATCTTCTGCTGGCTGGTTCGCGTTCGCCCTATTGAAGTGCCTCGTT---	6469
Qy	492	ArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerVal	511
Db	6468	-----CAAGAATPGCGCCCTCTCGTCGGATTTAT---GCAGGTGGCGCGCTCTGTT	6418
Qy	512	ArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn-----GlySer	527
Db	6417	CGTGGCTATGTTATCAGCAGTTAGGCCCCCAAGACCCATATAAGATGCGGTGGCGGA	6358
Qy	528	ValLeuProGluArgAlaLeuValGlySerLeuGluTyrGln-----	542
Db	6357	CTGTGCGAAGAGAG-----CTTCGTTTCGAAGTCGCTTACCGTTTGTGTAATTTGTT	6304
Qy	543	--LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla	561
Db	6303	CTTGTGCTTTCATTGAT-----GCAGGGCAGGTATATGAA-----GACGCTATC	6259
Qy	562	AlaAsnPhelysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer	581
Db	6258	CCAACTTTCAT-----AATTAAAGATTGCTACTGTTGGGTGCTCGTTACTATACC	6205
Qy	582	ProLeuAlaProPheSerPheAsp-----IleAlaTyrGlyHisSerAspLys	597
Db	6204	GCTTTTGGTCTTTTTCGCGATTGATCTTCGACACCTATAGCTGGCAGCGTGGAGAATCT	6145
Qy	598	LysIleArgTrpHisIleSerLeuGlyThrArgPhe	609
Db	6144	CGTATTTCAGTATATATTCGATCGCTCAGGCTTTC	6109
RESULT	30		
LOCUS	AY010120	20310 bp DNA linear	BCT 13-JUN-2002
DEFINITION	Xanthomonas oryzae pv. oryzae xanthomonadin biosynthetic gene cluster, partial sequence.		
ACCESSION	AY010120		
VERSION	AY010120.2	GI:15055553	
KEYWORDS	Xanthomonas oryzae pv. oryzae.		
SOURCE	Xanthomonas oryzae pv. oryzae		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; xanthomonas group; Xanthomonas.		
REFERENCE	1 (bases 1 to 20310)		
AUTHORS	Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.		
TITLE	Genetic locus encoding functions involved in biosynthesis and outer membrane localization of xanthomonadin in Xanthomonas oryzae pv. oryzae		
JOURNAL	J. Bacteriol. 184 (13), 3539-3548 (2002)		
MEDLINE	22053219		
PUBMED	12057948		
REFERENCE	2 (bases 1 to 15118)		
AUTHORS	Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-OCT-2000) C.C.M.B., Uppal Road, Hyderabad, A.P. 500 007, India		
REFERENCE	3 (bases 1 to 20310)		
AUTHORS	Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-AUG-2001) C.C.M.B., Habshiguda, Hyderabad, A.P. 500 007, India		
REMARK	Amino acid sequence update by submitter		
REFERENCE	4 (bases 1 to 20310)		
AUTHORS	Goel,A.K., Rajagopal,L., Nagesh,N. and Sonti,R.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAR-2002) C.C.M.B., Habshiguda, Hyderabad, A.P. 500 007, India		
REMARK	Sequence update by submitter		

```
COMMENT      On Aug 1, 2001 this sequence version replaced gi:11693112.
FEATURES
  source
    1..20310
    /organism="Xanthomonas oryzae pv. oryzae"
    /db_xref="taxon:64187"
    /note="pathovar: Oryzae"
    218..1999
    /note="Orf1; similar to YTFM from Escherichia coli"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="putative outer membrane protein"
    /protein_id="AAG38834.1"
    /db_xref="GI:11693113"
    /translation="MLKVFALFLCTLEVPLOQAQAQATIDKIEIKGLDGDGDAEMIE
    NIQVSLVAVKGECELEVLAAQEROTREALPEFYYSPTITLAAAPAGDKVT
    VITVDGEVRYVTSITGWAEODRYLGDLQKQEPREGQVFSHPQYDYEAFKVRIT
    RLAEQYFQADQTRVAITRAEHAADLDLMDSGRYDKGKVRFDYDREGLFPD
    LYWDEGSYTHEKDLRESLTKLDYFTIDIQKPEADQDQGVYDVKLTAKRKT
    VITAGLSGSEAGVGGVYVWARGHKMDTQLDYAQRKSLTTSYRIPAPRWLD
    GWYTSARLYDEQTDYIDLNVKLTGSRSGQINERMSAIASTINLRERWESGDDFE
    GAVYETSLYPOLOANYINVDRLFPRSGVSGOMFIRGGAEGAGSDTNFQGVYQGLR
    WFLAGDONSRLILRGEGGTTWTSDLVAMPSPSRFFAGGSNSIRGYAFREVGPTRPKD
    KFLKAGNVYTSAAEYEHYLNQPGWGGAVFVDSGSAFDDRPDWHGTGIGFGLWRSPVG
    PVRVDIAHLNDPDAQPOLYIDIGANI"
    2299..5853
    /note="Orf2; similar to YTFM from Escherichia coli"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="putative periplasmic protein"
    /protein_id="AAG38835.1"
    /db_xref="GI:11693114"
    /translation="MKPDLRLLGLRTLQLNAVQVSNATLNLGKSEEPETLPSWPESIP
    QITVPLAIOADKIDVONLRTLOQPMIVLHKVQGLQVATGALRDLRLDITDMGDF
    RLHGYVPSDDYHADUTATAVIPAARGTSPASGLVARGDLDQMEVATAGRAPAPLQA
    SLVSGDDQTNFAKVAYTEALDTSLLIPAAEQGANPPATPAINLQAAGSGNADLHG
    11KGAFSTQPSHIALDQVLTVEPLIDTFEGTQLRGTADLRDQTNQPSFPAVN
    ASGRLTPAADPATPOAPLVPVELDKARLVAGTALKAAATGRATVRDQQAELVFD
    SRGNDQKQKQVQAKTPGCTLDLSCVAPAWPELQWVNAQLKAFDPGYFAPGWNQNL
    SKGTASKGQOLPAPAGVSPGFEATAEVPSTQLQRALSAKNGKFLARBOGSGELO
    LALGNLRMDYSAQSLSLGRPLPWRGSDGLALQCTATEAGVVLDSVRVQARGAVEAL
    RLADIANSASVALQGDVPRNCRWQGVATLRIAPSKGDAAHALRQAPQSTQGAFF
    TLSDTCLGATGALCASANNFRAGVYVHGDAPLSLVQPMPLKQDQKQITLARGELSL
    DGSTKPGNAGWSRLTASPEGIRUGETRYAAVAGNPNRGELLRYDQFSNQADFTQQ
    QIOGKLGGFOGAGFVDAKFTNGWDAYAPLNGELYLNNSRLYWELEIADYAVRPEGRV
    EGHVSLGTRDKPLGGDATLSDFTAEPDSMGIALSEKGRFDALPDGSAKITASAKS
    GPGTLTVGGLSWFGTSTPLLNIRGDNVLAINTSELRIIANPQWQFQITDNTMHLRG
    KVTVPADIDLERLDRTSVSEDEVVVDVDPDEQTPASPLDNDLAIVLGDKVNMGTFG
    LKGLSGOMIARAPGREMANGLDVRGKYKAYGQDILTITRGQLTWNHNIVSDPRYS
    LRAERKTDITAGIDYSGRAESPADWSEPAKMSQSEALSVLGRGLSTASSDETQQ
    VSASAAISAGSLSIASQICATGLDDEAGVSQSSITGVSVVGFGYLSPLRVITGVGSM
    IGGSVUTLKLILSRGFDIEAESVIVETKGSVNWRRK"
    complement(5988..6934)
    /note="tnpXo; nonfunctional putative transposase due to
    frameshift mutation"
    complement(7188..7625)
    /note="Orf3"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAG38836.2"
    /db_xref="GI:19483794"
    /translation="MOCROAIATPLPHQSGMCLWEDVIEWDAQRIVLRSIAHRSDAHP
    LRSNGRLALHCEYGAQAMAVHGGLLGRASGKPYRPGMLVALRGVELHVHTLETLE
    AIDCEAVLMQGEDSQOYSFRLTHGQELLAEGRATVMLGAQA"
    complement(7612..10035)
    /function="probably involved in transport of xanthomonadin
    to the outer membrane"
    /note="Orf4"
    /codon_start=1
    /transl_table=11
```

```

    /evidence=not_experimental
    /transl_table=11
    /product="putative inner membrane protein"
    /protein_id="AAG38837.1"
    /db_xref="GI:11693116"
    /translation="MPRLPACVRIQQDTHMALELNANRRRIGLALLALLAVAGF
    WLSWTLKVTGDLRKFMPAPRTPAOKLTIELGEGPCSLLLMALNSDPPTLEAQSOQ
    /transl_table=11
    /product="putative inner membrane protein"
    /protein_id="AAG38838.1"
    /db_xref="GI:11693117"
    /translation="MPMCRITLPMMLILLFSAAPLCAAPPETLDVGQVLERLARPAVYS
    TEFVLGRSSILKTPYRVEGHYRRPDAQTLVREVSAPYHETTTITGDEGVLEREGKSP
    RRFSLRVPELAGLKGFGALLSGDRALLEOHYSVSGGQPOAMLTLQPKDAAVAKQ
    VRQLRYGRKDELRCIESYPAKGDVQRTLLACAAREAAAGVTDATATJTLCHGSCA"
    complement(10691..11665)
    /note="Orf6"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="putative acyl transferase"
    /protein_id="AAG38839.1"
    /db_xref="GI:11693118"
    /translation="MSKHMKRQEGGGGFALMLIRGIARHGRVYVGRALLPYTLVFL
    LVRAPEAASRAYLAVLGRPATLRDVARHIHTFASTILDRVYMLCGOMORFVLDJSG
    LDLOLTHRGGRGVLLFGSHLSEDLARVLATERPDQVQVVLQKANHNRDILGLAL
    NPQLAANITDAGMDATSIYMAKEATDAGALLVALLIDRREGDPALPAMFLGRNMFPP
    TSPMLTAAALKVPVLAFLYHGGNHIELAFETTSSEGDVPHRQAPMLAVLMRDYAA
    RLEHTCYAPYNNMFYDFWNTHHADVPHPADVDATVQVRRAMRRTA"
    complement(11662..11949)
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="putative dehydratase"
    /protein_id="AAG38840.1"
    /db_xref="GI:11693119"
    /translation="MDPVPHDHPCLPGHPGPRVVPVGVVLDHVLAQVAELHGPRAA
    ARLPQVAFVQPLPGQTASVTLGEGDQPRWRFRVQADVVLVSGELVABAGT"
    complement(11961..12710)
    /note="Orf8"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="putative acetyl transferase"
    /protein_id="AAG38841.1"
    /db_xref="GI:11693120"
    /translation="MPLARMPSLAWAFNLLQLAFTLAFTAGGIVYVALALMAITRDA
    DRMLRGMWMAVPLFHHGACVTIVBGRERVDKSNLYFVSNHOSIIDICALFYALPV
    DRLLKKEEMKVPFVNYWYATGMFLDRDRHRRAGALVRRQVAALLRQGLCLFPE
    GTSRRTCALLPFKVLQALADAGVQVVPVLDGCGRVLVVDGLFRVRRPGIIRVRICT
    PIATVGPDAQDRQOLTEQAHKAVAMLHPRI"
    complement(12764..13438)
    /note="Orf9"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
```

```
/product="hypothetical protein"
/db_xref="GI:19483795"
/translation="MGVLLAVAYSPLAHWANASHREVAIVAGLLVLMVLIEMVAR
RAMAALLALLAGLVGWHSPYAMVLAAPVFTGVWAMFFGRSHRGRTPLITRI
VEGLYGAGPVPPEQRRTTRLLTAWALLLGLGLLANLVGLCAEPSPGVLTOLGYS
PLPTSDARSLFANLLCYGVIGFFTEYLLRGRWFPORPYRNLPDFLQMARLGPDE
WRDLRL"
CDS
complement(13485..13883)
/note="Orf10"
/codon_start=1
/evidence-not_experimental
/translation="MGVLLAVAYSPLAHWANASHREVAIVAGLLVLMVLIEMVAR
RAMAALLALLAGLVGWHSPYAMVLAAPVFTGVWAMFFGRSHRGRTPLITRI
VEGLYGAGPVPPEQRRTTRLLTAWALLLGLGLLANLVGLCAEPSPGVLTOLGYS
PLPTSDARSLFANLLCYGVIGFFTEYLLRGRWFPORPYRNLPDFLQMARLGPDE
WRDLRL"
/translation="MLGLAVTATYTTADATPHSDGRPHCCCLFLLPPYPDLWKSWM
SSQTAARELAELLVSLNLEVDQPADIDPEAPLENTGLDLSDALELALAIKRVG
FQLRSDNDNRIFASLRALSRAVENKTV"
```

Alignment Scores:

Pred. No.:	1.96e-09	Length:	20310
Score:	292.00	Matches:	151
Percent Similarity:	40.43%	Conservative:	92
Best Local Similarity:	25.12%	Mismatches:	261
Query Match:	9.20%	Indels:	98
DB:	1	Gaps:	27

US-09-857-669-2 (1-609) x AY010120 (1-20310)

QY	58	AspThrGlnAspSerGluIleuLeuAspMetValGluGluHisLeuProLeuIleThrGln	77
DB	323	GAGCGGACGCGCGAGATGATCGAGACATCCAGGTCTCTTTGCTCTCTAT	376
QY	78	GlnGlnGluGluValLeuAspLysGluGln	93
DB	377	-----GAGCTGTGGCAAGAAACAGGCGAGTCGCGTGTGGAGTATCTGTGGCA	427
QY	94	GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal	113
DB	428	CAGCGGACGCGCAGCGCGGCGGTGGAGCGGTTGGCTATTACTCGCAACCATC	487
QY	114	SerLeuThr-----GluLysAspClyAlaTyrThrValHisIleThr-----ProGly	129
DB	488	ACGCTGGTGGCCCGACGCGCGGCGAGCAAGGTGACCGTGTGATCACGTTGGATCGTGA	547
QY	130	ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly	149
DB	548	GAGCGGTCGCGGTGCGCACCTCGCATATTTCGATCACCGGTCGGCGCAGCAGGAC	604
QY	150	AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpClnGlnProValGlySerAsp	169
DB	605	-----CGTATCTAGGCGAGGACCTCAAGCAGTTCGAGCGCGCGAGGGGCGAGTGC	655
QY	170	PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys	189
DB	656	TTGAGTCATCCGACGACGAGCAAGGTGCGCATACCCTGCGCGCTGCGCGAGCGT	715
QY	190	GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr	209
DB	716	GCTACTTCGATGCGCACTTATCCACGCGCGGTACGATCACGCGCGCGAGCAGCGG	775
QY	210	ValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly-----	225
DB	776	GCGGATATCGATCTCAATTGGACAGCGCGCGCTPACGACATGGGCAAGGTGCGCTTC	835
QY	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
DB	836	GACTACGATAC-----TTCCGCGATGGTCTGTTCACCGCGCTGCTGATC	880
QY	246	PheGlnProGlyThrProTyrAspLeuLeuLeuAspPheGlnGlnAlaLeuGlu	265
DB	881	TGGGACGAGGCGGACTACTACCAAGGCAAGCTGGATCGGCTACGCGAGTCGCTGACC	940

QY	266	GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu-----Gln	283
DB	941	AAGCTGGGATTACTTCAGCACCATCGATATCCAGCCCAAGCCGAAGAACCGACGACAG	1000
QY	284	GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr	303
DB	1001	GGG-----CGGTCGCGGTGAGCTCAAACTCACCGCGCCCAAGCGCGCTCTACACCGCA	1057
QY	304	GlyIleArgLeuAspSerSerGlyTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyr	321
DB	1058	GGCTTCAGCTACCGCAGCAGCGCGCGGTGTGCGCGGTGGCGTGGAGCGGCGTTAC	1117
QY	322	TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu	341
DB	1118	GTGAATGGG-----CGTGGCCACACAGATGGACAGCAGCTGGATTACGCGAGAACCGC	1171
QY	342	ThrThrLeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrp	359
DB	1172	AAGAGCTGCACCCAGCTACCGCATACCGCGCTTCGTTGGCTGGTGGCTGGTACACG	1231
QY	360	ThrSerAsnValSerTyrAsnArgSerThr-----ThrGlnAsnLeuGluLys	375
DB	1232	GCTTCGCGCGCGCTTTCAGCAGCAGCAGCGATTCATCGACCTGCGCAACGTCAAACATC	1291
QY	376	ArgAlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeu	395
DB	1292	ACCGCAGCGCGCAGCGCG-----CAGATCAACAGCGCTGGAGCGCGATCGCGTGCATC	1345
QY	396	GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsn	415
DB	1346	AAGCGC-----TTCCGCGAACGTTGGCGTTTCAGCAGCGCGCGCATTTTCGAAGCGCG	1399
QY	416	SerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHis	435
DB	1400	GTGTACGAGAC-----LeuSerSerThrAlaLeu-----TCCAGGTGTGATCTAT	1426
QY	436	ProGlu-----AsnGlyHisTyrLeuAspLysIleGly-ThrThrLeuGlyThrPhe--	453
DB	1427	CGCAGCTCGACGCCCACTAC-----ATCAATCTCGACGACCGCGCTGTTCG	1474
QY	454	-----LeuSerSerThrAlaLeu-----Tl	460
DB	1475	CGCAGCGCGGTGTCGGGCCAGATGTTCATCGTGGCTGGCGAGGTCGCGGCTCCGAT	1534
QY	460	eArgThrSerAlaArg-----AlaGlyTyrPhePheThrProGluAsnLys	475
DB	1535	ACCACTTCGCGCAGGTGTATGGCAGCTGCGCTGCTTCTCGGCG-CGTGGCGACAACAG	1593
QY	475	SlyLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAla	495
DB	1594	CGGTTG-----ATCTTGGCGCGGAAGCGCGCACCCCTGSGACCGACCGCTGT	1644
QY	495	aAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrG1	515
DB	1645	CGCGATGCGCGCAACGCTCGCTTTTCGCGCGGTGTTTCGAACAGCATTCGCGGTATGC	1704
QY	515	uLeuAspSerIleGly-----LeuAlaGlyProAsnGlySerValLeuProGluArgAl	533
DB	1705	GTTTCGCGAAGTCGGCGCGCGCGCCCAAGCCGACAACTTCGCGCTGGTGCCCAAGAA	1764
QY	533	aLeuLeuValGlySerLeuGluTyrGlnLeu-----PropheThrArgTh	548
DB	1765	CGTGGTACCGCGCATCGCGAATACGAGCACTACTCTCAACGGTGGCGCGTGG	1816
QY	548	rLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhelysArgMetLys	568
DB	1817	---GGCGCGCGGTGTTGTCGACAGCGCGCGGCTTCGACGAC-----CGCCCGGA	1866
QY	568	sLeuLysHisLysSerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPh	588
DB	1867	CTGGCACACCGGTATCGCTTCGCGCTCGCTGCGGTTCGCGGTGGCGCGCGGTGGCGCT	1926


```

gene      complement(4170..4550)
CDS       /gene="XAC4202"
          complement(4170..4550)
          /gene="XAC4202"
          /note="putative; ORF located using Glimmer/Genemark"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA39037.1"
          /db_xref="GI:21110633"
          /translation="MFQQLPACGRALNPLACADVDRILAEIHAAGDAEVELVLVLD
          GDPLSERCARALRAVDALPPTPIELTHADQEPWLAHQAPLAVVITPHDAPR
          AYAMSLGIAARCLPSLHAPLRYAA"
          complement(4859..8419)
          /gene="XAC4203"
          complement(4859..8419)
          /gene="XAC4203"
          /note="identified by sequence similarity; putative; ORF
          located using Blastx/Glimmer/Genemark"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA39038.1"
          /db_xref="GI:21110634"
          /translation="MYLPDLRLPLRLQLQDAQVSNATLNLGKSEPTLSPSES
          LPQINPLAIQADKIDVQNLRLITQOPMIVLHVKGGLDVATGELKTRLDVITDMG
          QRLHGDYLPNDIDRADLTATVLPAAGRTPASIGLVARGDLDKMEVATAGRAPADL
          DASLFTGRDDPTWAKVATEALDLSLIPAAEGANPPATPIALNLQASGKGNADL
          HSGLSQELSLAQPSHIALQDVLTVEPLVIDTFEGRTQLRGTDARFDQNFSPRFA
          VNASGRTFPADAPDPAPLVEQLQDARLGAVGTAKMAAIGRAVERDQGOAEIV
          FDSRGDKAQKQVAKTPGGSLDLTGEVAPALOWDVSQAOLAKTDPGCFAPAGNNG
          NLGSKTAGKROLPAAGCVSPGFETAIEPSLTGOLPQALSGNCFALRGQOGBCE
          LQIALGNSRINAKGVGDQDLIAQLOPLQDLDPGATGVVRQLOVSKRQADPIT
          ADIAGLRWDITSAQNLISRLGRLPWRSQDQALQATIEAGVLDLVSVRQARGAVE
          ALRDADIANMSVAQGDVRRNGQVATRLIAPAKGDWLSRQAPQSTQGA
          AFTLSDTCGAATGALCASANPREGVHGDALPLSLVQWLPKQEGROIYLRGEL
          SLDSGSPRGNAWEGSLRIASPEGGIRLGETRYAAVAGNPNRGLLRDQFSVQADFT
          QOQIQGLGFGGACGDAKFNWDYAPLNGELYLNMSRLYLWELVADLVVRPKG
          LVEGHSRLGTRQKPLGGDATLSDFTAEPYPMGLTSEGRFDALPDGSAKITASA
          KSGPTTVLPGGLSFRCTSTPILLNLRGDNLVAYNTSELRIANPDWQFQITQNTWL
          RGVTVPEADIDILERLDRTSVSDVVLDPDPEOTPASPLMDLAIVLGDKVNSG
          FGLKGLSGQKQIARPREGHTANGLDVRGRYKATGQDILITRGQITWNNIVSDPR
          VSLRKEICGVTAGIDVSGRAESPADVWSEPMQSSEALSYLVLRGLSTASSDET
          QVSAASAAKLSAGSUSIAQIAGLGLDEAGVSSSTLGSVVGFKYLSPLXYGVGV
          SMWGGSVLPYLLYLRGFDIEAESSTIETKGSVNNRREK"
          complement(8710..10491)
          /gene="XAC4204"
          complement(8710..10491)
          /gene="XAC4204"
          /note="identified by sequence similarity; putative; ORF
          located using Blastx/Glimmer/Genemark"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA39039.1"
          /db_xref="GI:21110635"
          /translation="MLKVAFLFLCTLPVLOAVARATIDKIEIKGLDGDQDAEMIE
          NIOYSLYFAVCKEGRLEWLLAQAPROTREALPEPGYKSTPIILAPRAGDKVT
          VYIVDGEFVRVTRSHISITWAEQDRYLQQLQKQEPREGQVFSHPQERKGLFT
          RRLAREYFQADTQRVAITRAEHAADIDLNWDSGRRYDMGKVRFDYDFRGLRDP
          LVYWDGSGYHEKGLRRESLTLDYFSTIDIOKPEEADDQGVDPVKLTFRKRT
          VYTSGLSYSGESGAGRGVGVYRYNARGHKMDQLDYAQRKSLTTSYRVPFRWLD
          GWYTPASARLYDEQTEYIDLNRKLTGSRSGQINENWSAIIINALRERWRFSSGDDE
          GAVYETSTLYVLOQANYVNVDDLPPRSVGSQAQFIRGGAEAGSDTNFQQLYQLR

```

Alignment Scores:

```

Pred. No.:      9,28e-10      Length:      10839
Score:          291.50      Matches:      149
Percent Similarity: 37.32%      Conservat:      85
Best Local Similarity: 23.76%      Mismatches:    245
Query Match:      9.19%      Indels:      148
DB:              1          Gaps:      24

```

```

US-09-857-669-2 (1-609) x AE012071 (1-10839)
QY 58 AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGln 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10386 GACGGCGAGACGACGCGGAGATGATCGAGACATCCAGGTCCTCCTCTGTAT----- 10333

QY 78 GlnGlnGluGluValLeuAspLysGluGln-----ThrGlyPheLeuAlaGlu 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10332 -----GAGCGCGTGGCAGAGGACGCGGAGTCCGCTCTGGAATATCTGTGGCA 10282

QY 94 GluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10281 CAGGCGGAGCGGACGCGGTGAGCGCTGTGAGCGCTTGGTCTATTACTCGCGCACCAT 10222

QY 114 SerLeuThr-----GluLysAspGlyAlaTyrThrValHisIleThr-----ProGly 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10221 AGCGTGGCGGCACACCGCGCGGCAAAAGTGCACGCTGGTGATCAGCGTGATCGTGGC 10162

QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10161 GAGCGCGTGGCGGTCGCACCTCGCATATATCTCCATCAGCGGTGGCGGAGCAGGAC--- 10105

QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAsp 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10104 -----CGCTATCTGGCGCAGGATCTCAAGCAGTTCGAGCGCGCGGAGCGCAGGTC 10054

QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10053 TTCACGCCATCGCGCATAGCAAGCAGCAGGTCGCGATCACCCTGCGCGTCTGGCGCAGGT 9994

QY 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9993 GGCTACTTCGTGCGGACTTCCACCGCGCGCGTTCGATCAGCGTCCGCGGATCGG 9934

QY 210 ValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly----- 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9933 GCGGACATCGATCGAATTGGAGCAGCGCGCGTACGACATCGGCAAGGTGGCGTTC 9874

QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9873 GACTACGACTAC-----TTCGGCGATGGCGTGTTCACCCCGTGGTGTAC 9829

QY 246 PheGlnProGlyThrProTyrAspLeuLeuLeuAspPheGlnIleAlaLeuGlu 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9828 TGGGACGAAGCAGCATCTACACCAAGGCAAGCTTTCACCGATTCGCGCAGTCTGCTGACC 9769

QY 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu-----Gln 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9768 AGCTGGATTACTTCAGCACCATCGATCCACCCCAAGCCGAGAGCGGCGGACCCAG 9709

QY 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9708 GGC---TGGTGGCGGTGGAGCTCAAACTCACCGCGCCGACGCGTCTACACCTCC 9652

QY 304 GlyIleArgLeuAspSerGluTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyr 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9651 GGCTTCAGCTACGGCAGGCAAGCGCGCGGTGTGCGGGTGGCGTGGCGGGCGGTAC 9592

QY 322 TyrAsnLeuPheAsnLysGlyTyrIleCysValThrValTrpAspMetAspLysTyrGlu 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9591 GTGAACGGG-----CGCGCGCACCAAGATGGACGCGCGTGGATTACCGCGGAGAACCGC 9538

QY 342 ThrThrLeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrp 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9537 AGAGCTGACCACCATGTTACCGCGTGGCGGCTTCGTTGGCTGGATGGCTGTACACC 9478

QY 360 ThrSer-----AsnValSerTyr 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9477 GCCTCGCGCGCGCTGTAGCAGCAGCAGCAGCAGTACATCGACCTGCGGCAACGTCAAATC 9418

QY 366 AsnArgSerThrThrGlnAsnLeuGluLysArg----- 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9417 ACCGGCAGCGCGGCGCAGATCAACGAGCGGTGGAGCGGATCGCTCGATCAACGCC 9358

```

```

Oy 377 -----AlaPheSerGlyGlyIleTrp----- 383
Db 9357 TTGCGGACGGTTGGCGTTTCAGACGGCGACCATTTGAGAGCGCGGTGTACGAGACC 9298
Oy 384 -----TyrValArg----- 386
Db 9297 TCCACACTGATCATCCGCAACTGCAGGCCAATACGTCAATGTCGACGACCGCGCTGTT 9238
Oy 387 AspArgAlaGlyLeuAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 9237 CCGCGAGTGGCGTGTGCGCGCAGATG-----TTCATTCCGCGCGCGCGCGAGGT 9187
Oy 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLys 426
Db 9186 CCGCGCTCCGATACCAACTTCGCCAG----- 9160
Oy 427 ArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle 446
Db 9159 ----- 9145
Oy 447 GlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAla 466
Db 9144 CGCTGGTTTCTGGC----- 9127
Oy 467 GlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla 486
Db 9126 GGC-----GACAA-----GCCCGTGTGATCTCCCGCGGAGGC 9091
Oy 487 GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
Db 9090 GGCACCACTGGACACGACCGTGTGGCGATCGCGCGAGCGCTGCTTCTTCGCGCGT 9031
Oy 507 GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly-----LeuAlaGlyPro 524
Db 9030 GCGCCCAACAGCATTCGCGGTATCGCTTCGCCGAAGTGGCCACGCGCCCAAGCGG 8971
Oy 525 AsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlyTyrGln--LeuP 544
Db 8970 GAGCAATTCGCGTGGTGCCCAAGAACGTGTCACCGCCAGTGCCTGCAATACGACATTAC 8911
Oy 544 roPheThrArgThrLeu-SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsn 563
Db 8910 CTCAGGCGCGCCCTGGCGGCGCGGTGTCGTGATGGCGCAGCGCTTCGATGAC 8851
Oy 564 PheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeu 583
Db 8850 -----CGCGCGGACTGGCACACCGGTATCGGCTTCGCGCGCTGCGCGGTG 8797
Oy 584 AlaProPheSerPheAspIleAlaTyrGly-----HisSerAspLysLysIleArgTrp 601
Db 8796 GCGCGCGTGGCGTGATATCGCATCGCTTGACACCGCCGACTCGCAATTCAGCTC 8737
Oy 602 HisIleSerLeuGly 606
Db 8736 TACATCGATATCGG 8722

RESULT 32
AE012531/c
LOCUS AE012531 11870 bp DNA linear BCT 23-MAY-2002
DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 439
of 460 of the complete genome.
ACCESSION AE012531 AE008922
VERSION AE012531.1 GI:21115353
KEYWORDS
SOURCE Xanthomonas campestris pv. campestris str. ATCC 33913.
ORGANISM Xanthomonas campestris pv. campestris str. ATCC 33913
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xanthomonas.
REFERENCE 1 (bases 1 to 11870)
AUTHORS da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,

```

```

Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Clapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Curso-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.A., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,J.H.A.,
Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (5687), 459-463 (2002)
22022145
12024217
2 (bases 1 to 11870)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Clapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Curso-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,J.H.A.,
Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitajima,J.P.
Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil
FEATURES
Location/Qualifiers
1..11870
/organism="Xanthomonas campestris pv. campestris str. ATCC
33913"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
480..2099
/gene="XCC4080"
480..2099
/gene="XCC4080"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="polyvinylalcohol dehydrogenase"
/protein_id="AA043301.1"
/db_xref="GI:21115354"
/transl_table="MSLDTAFRAASGQINRSFSGNLPASPNNPNTAGGCGYNSRA
PCWKINRTTAPLTKPAWTLTTCIDISATPSVEGGTLVYPDWGCTLYSDVDSQSKPW
QAKLSVDYTGANSVTRNTPAISGSIIVGDAQSGVTMAIDKRTGLKLMSSVAESNPQA
RTSSPYNGRIYVIGSSGDCWGLAANHTFSFGSVALDLNTGKKVQYVYNYPEGY
TCASVGOVAIDPISRLRYFTGNYNIPAGVNCVSNARSGTSLVTODELNCMAP
DNVDSVALNLDGTGLANADRAQYDAWNLSCIVAPTNGLCPNTAFNLTPDYDFG
GAGVQLFTVWRDGRKPOLQVAGCGSGIYWFADARTGAQVSTGVCPTGGVEMGTG
FDPYKSOVYVALINSSNATYTLGPDNTVSHNGSSNAALDAATGKIKQOVKYPGVNSYV
PNVPAGGQSGISTVSPNLLYAGSWAGNVALDPTGTATLNTYNTATGSSVLTSSPAVVDGAL
YWGAGYGRFNFEGTAATSNQLIKFVPDSGAAR"
complement(2282..2587)
/gene="XCC4081"
complement(2282..2587)
/gene="XCC4081"
/note="putative; ORF located using Glimmer/Genemark"

```


Db 8864 CGTATCTGGCCAGGACCTCAAGCAGTTTCGAGCGCGCGAAGGCGAGGCTCTTCAGCCAT 8805
Qy 173 AspSerTrpGluAsnSerIysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro 192
Db 8804 CCACAGTACCAAGCCAGCAGAGTGCATCACCOCGCCCTGGCGCAACGCGGTATTTTC 8745
Qy 193 LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeu 212
Db 8744 GATCCGCACTTCCAGCGCGCGCGGTGGCGGTGACCGCTGCCGAGCATGCCCGCGACATC 8685
Qy 213 AsnValValAspSerGlyArgProIleAlaPheGly-----AspPheGlu 228
Db 8694 GATCTGAACCTGGACAGCGCGCCCTACGACATGGGACCGCTGCCTTTGACTAGCAT 8625
Qy 229 IleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnPro 248
Db 8624 TAT-----TTCCGCGCGCGGCTGTCGAGCGCGCTGGTATTGGAGCAG 8580
Qy 249 GlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGly 268
Db 8579 GCGAGCTACTACCAAGCAAGCGTGGATCGGCTGCGGAGTCACTACCAACGCTCGAT 8520
Qy 269 HisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu-----GlnGlyAspArg 286
Db 8519 TACTTACGACCATCGACATCCAGCCCAAGCGGAGGAGCGGATGACCAAGGCG---CGG 8463
Qy 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg 306
Db 8462 GTCCCGGTGGACGTCAAACTCAACCGCGCAAGCGCACCGGTATACCTCCGGTCTGAGC 8403
Qy 307 LeuAspSerGlyTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyrAsnLeu 324
Db 8402 TATGGCAGCAGAGCGCGCGGTGTGCGCGGGCGGTGGAGCGCGCTATGTCAAT--- 8346
Qy 325 PheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyrGluThrThrLeu 344
Db 8345 ---TCGCGCGCGCACAGATGACAGCGAGCTGGACTACGGCGCAACCGCAAGCGGCTG 8289
Qy 345 AlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
Db 8288 ACCACGACTACCGGTGCGCGGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8229
Qy 363 ValSerTyrAsnArgSerThr-----ThrGlnAsnLeuGluLysArgAlaPhe 378
Db 8228 CCCTGTACGACGAGCAGCAGCAAGTACATCGACCTGCGCAGCGTCAAACTACCGCGGAGC 8169
Qy 379 SerGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGlu 398
Db 8168 CGCAGTGGC-----CAGATCAACGAGCGCTGGAGTGGGATCGCTCGATCAACGCC--- 8118
Qy 399 PheLeuAlaGluGlyArgLysIleProGlySerAspIle---AspLeuGlyAsnSerHis 417
Db 8117 TTGGCGGAGGCTGGCGGTTTACAGCATGGCAGCGATTTACCGATCGCGTGTACGAGACC 8058
Qy 418 AlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn-----AsnValLeuHis 435
Db 8057 TCCACGTGTATCTATCCGCGAGCTGCGAGTGCAGCACTACGTCAACGTGATGACCGGCTGTC 7998
Qy 436 ProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSer 455
Db 7997 CCGCGCAGCGGC-----GTGTCC 7980
Qy 456 SerThrAlaLeuIleArgThrSerAlaArg-----AlaGly----- 467
Db 7979 GCGGAGATGTTTATCCG 7920
Qy 468 -----TyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 7919 CTCTACGCCAGCTACGGTGTCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7869
Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 7868 ATCTGCGTGGCGAAGGTGGCACCACCTGGACCGCATCTGGTGGCGATCCCGCGGAGC 7809

Qy 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 7808 CTGCGGTTCCTTTCGCGTGGCCCAACACGATCCCGCGGTATGCAATCCCGGAAGTGGC 7749
Qy 521 -----LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer 538
Db 7748 CCGCGCAGCCCAAGCGGAGGAATTCGCGCTGGTGGTCCCAAGACGTGGTCCACCGCCAGC 7689
Qy 539 LeuGluTyrGlnLeu-----ProPheThrArgThrLeuSerGlyAlaVal 553
Db 7688 GCCGAATACGACGATTAACCTCAAGCGCGCGCGCTGG-----GGCGGCGCGGTG 7641
Qy 554 PheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeuLysHisGlySer 573
Db 7640 TTCTGTCATAGCGCGGAGTGGCTTCGATGACACCCCGGACTGGCGCAC-----GGCAT 7587
Qy 574 GlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGly 593
Db 7586 GGCTTCGGCTTGGCTGGCGCTCGCGGTGGGCGGAGTGGCGGTGGATATTTCGCGACGGC 7527
Qy 594 -----HisSerAspLysLysIleArgTrpHisIleSerLeuGly 606
Db 7526 TTGAACGACCCGATTCGCGAGTTCACGCTCTACATCGACATCGGG 7482
RESULT 33
AF003588 334520 bp DNA linear BCT 28-NOV-2001
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section B/19.
ACCESSION AF003588 BA000019
VERSION AP003588.1 GI:17131110
KEYWORDS
SOURCE Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
1
AUTHORS Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
Yasuda,M. and Tabata,S.
Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120
DNA res. 8 (5), 205-213 (2001)
2 (bases 1 to 334520)
DIRECT SUBMISSION
Kaneko,T.
SUBMITTED (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/cyanobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
Location/Qualifiers
1. 334520
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/note="synonym:Anabaena sp. PCC7120"
complement(761..3619)
/gene="all2020"
complement(761..3619)
/note="all2020"
/note="ORF_ID:all2020"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="BA073719.1"
/db_xref="GI:17131111"

FEATURES
sourcegene
CDS

EQEAQELTRLOERQNRVTVTAALGELSLTNSQGDFFLEGTPLLVAAVCNAM
GLKISPLRSNLRQRYVEPLEAIVRASIRMRVLRDNWMLKAGALVATKNOPV
ALLPAGNRIEIPDNVSRVYVNEGVAATLAPVAYIRSLPDKADALVIAIVKFTLRG
HWDMLQLFTGIFITILGMITPVATSIIDNAIPDSRSLTQIGLGLVSAAGAGL
FQAQGLAFMRMETSGDAQOAGVMDRLNLPVFFREYTTGDLISRISSVAIRROL
SRTLINEFITFALFYLGQFYNYKALIAVASAVAIAITISGILLSSKVRPLI
EQVGNIFQOTVOLINGSLKLIHAGAEERAFAMSKNTYIKLELSTOLLSEFVVIEN
TIIPIQSPVLFWFTKLLBEAOTSGSIAPSLGTFIAFNATFNTRGTGNTLSNTVTO
LIQILPQORQTOPILTIPEVNLKSPDGPRLAGKITVDNATFCYRDRDGNLLDNVSI
APEGFIALVSSSGSKTLLRLLGRETPOAGSIHYDQDLSGLDNDVNRQLQGLV
QNGQSLASITENAGDAQTLDENAEARMSGFAEDVTAMPQMOMHTVVSEGGNLSG
QNGQLLARALKPRILLFDENATSDALNRTQAIIVSESLDKLQVTRIAHRLSTIR
NAHRIYVLAGRVVQOQTEFELANTEGIFIAQLMARQA

gene

CDS

/gene="all2021"
complement(3819. .5957)
/gene="all2021"
/note="ORF_ID:all2021"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein"
/protein_id="BAB73720.1"
/db_xref="GI:17131112"
/translation="MKPTLLQMEAVECGAALGIILGVYGRVVLPLPOLRRDGVSRD
GSKASNMVKAARTYGLEAKFKKELKQVQQLRPPIYVFNPNHFVYVVEGFGKHQYIN
DATGPRHVSQEFDEGTYGIVLMEPGAQFQGGRRPSMIASLWSRQGAESALVYC
LVAGFLFTIGLIVPVSQIPVDVILVOGRMLHLLPMLAMAIAALQSGSLTLRLRY
IRLKLRLVAGMSRFLHILRLPVSYAQRFAEETSNRTLLNDQVADYAGOLATTI
IDTMLVYALVMQYDMVLTLMVSPAAVNVLTQWISQRIDANORLIQYGAAG
ISIAITQSLETALKSGLESDFSRWSGYTKALNSQOELGITNOTFVSLPTLSALSA
MULLIVGPRVMDKLSIGMLIAFOGLMGFQOQPVNVNFGTILQELSGNLRDDVA
LNPVGEVGGKTSISSLOALPKLOGYVELKNITFGYSRLEAPLLENLISIKPGORVA
LVGSGSGKSTIAKLVSLGVQWPWTELYFDSQPREISLQILNLSVANVQEQIILFGG
TYRDLNTLWDTYFDKLNMRACQAAITADVFSMGYDAELIEGAALSGGQRRL
IARALNPSLITFMDNLSALDAETEKITIDONLRRGCTCIIVAHRLSTIRDCQIIV
LERGVQOQTHQELMQVEGAYSLRITG"

gene

CDS

/gene="all2022"
complement(6028. .7395)
/gene="all2022"
/note="ORF_ID:all2022"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73721.1"
/db_xref="GI:17131113"
/translation="MQVSPKSLPLVALGSIYGVAFINSIYGISPIITVEGRGVLIYP
SNVPLQSKAGQIMALNVKVDVKKGVQVLTIDQAEKRLKQOQKGLTQLESQDK
AVGSLQGRLEQEPESQOQRYLRQRIELEALTPLAKTNTSDQOQROGLQRLQ
QQAALTAVFLRMKTRQELFOKEGALSDEALKAEQYIIONLEKADIQAQKELDVK
EQQERDRENLTIVADQSQKOLDQSQASVQADLENYARKKEIQVKKREIGSLG
LQNSQSIISQHSQRILEITLTTPGVNAGTLATIEANPQSKVLGYTFYFVAGG
KIQPMQSQITPTQVTKRERFGVIGVNTSISRPTITKEAANEVGNSEVLEGVLSQOQ
GLIQVFSHLELDANTPSGYKWSSTGPHLNISSGTTTVTVRVKVEERAPITFVLPILRS
ASGIY"
complement(7472. .9502)
/gene="all2023"
complement(7472. .9502)
/note="ORF_ID:all2023"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73722.1"
/db_xref="GI:17131114"
/translation="MSDVCLVMPYAAVERPSTALGILLKARLQOAKLEPTILYLPNIOF
ABEIGLDVAMISENFIEFTLGMFTFSGVLFPFNPNSHSEYFDLISPCRSYENIPLD
IINKZFAFVRCAGATAFIESVAQSVANLQPRIIGCSSTFOQHCASLALLKRVRELAP
RITTFGGANGECAMLATHTHSWDFIISGEGDOLFNEFCYQLDGGREVDPISILPY
GVISPSHTSTKISTAAPRASVQNDQVPIDPDYDFTQTLHASTLSPYIQPGLLIETS
RCGMWQKHCTFCGLNGSGINYSKSPDRVLKEFTOLCERYGLKRFQVVDNIDMNH
MNTVIVPLAAGDEAYNIYEYKANLRWEHLQOLAQAGVRCIQPGIESMNSILKLMNK
GNTANMNVQLKWARELGIIIVFNILADFPGESDENYENKMWLPIIAHLQPPSGNSE

gene

CDS

TRYDNFSPYERPSPDWGLNLSPLRPYKYVYPLAPEILSELAYFFEDQNNLVTTRYATEE
NQOTSEKTERGLENLKLWNOMNRQWSLYTKYEHKPEYLPILSHMDSGLITTYDSR
PCATSESLPLGLELAYWYTKCDRALTYREVSNKRKLDVPVSWDEIFPIVSELREKR
ILLDIDGRLLSLALKGSCSPYLSSELPAGYIDLHEFRHRNSKSSSEKVLPPN"
/gene="asl12024"
complement(9616. .9909)
/gene="asl12024"
complement(9616. .9909)
/note="ORF_ID:asl12024"
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73723.1"
/db_xref="GI:17131115"
/translation="MSLVMTWKELEEKIIVKAGODESFKQALVSNPRSTLEKEGINLP
SSIDVNLVETTPGNLSLQALPNSFODTSLSEAELESVAGGRISGEVSVTVKVF"
complement(10195. .13482)
/gene="all2025"
complement(10195. .13482)
/gene="all2025"
/note="ORF_ID:all2025"
similar to Mrsd protein"
/codon_start=1
/transl_table=11
/protein_id="BAB73724.1"
/db_xref="GI:17131116"
/translation="MNVSKQSLIKIIBRSNISIERSLSFIFNDROIINANNNAKLEK
WFOCTAKQGNHFERKRLDGLDNTISCVIGNVNADEKYLPNVETYLKAGQVATF
VIPENSKNFQOQRYLDPERPFPFETILVPFVDAKKLIESAGSSYLLSQNAHLIQ
MERSLLEHLTGICISPLEFESIFRSLSKSPVRFVFNQIASCNSKOQYTFVYNKLLKD
GLLSFQRYSVLARIATSDVFEWATGEFLRSIADSKDIQITPOEOKLGOVYSVQ
WAPTRERRMQQAIIAQOQQAIRGKNSVTWIGVYLSCLTNQSLNPLEPNLSYMD
INKFISSSSYDDVIAHLRQMDADLAKIISIRGSLYSCLTNQSLNPLEPNLSYMD
GGCGLSLAALASVTRNFEGLSLASLOSRLTKLDPHIDIQOKLIQMGIGGATG
LASIVYALVRUSEFGEVLQDAQALASLWINEATVQAGVMGVGTILSLAL
YKATKEPATLQOASIGNYLLDTRVTVDEVYQANTNAGKVLGLFSGHGAAGTAYALLQ
LIAATDEPIFASAREATLEQRLFSNCCQDIDINSLSLSTTTCWCSGLIGIS
VARLGTNLITDTEIRQELDNLTHSSNQCIDVNICCGNFSHIEMNLVASKJSHP
ELLETANNRATOVLSRANOIGYFOLLPHLPADVTSFGFCOGIAGIGYELLRLSCPDLL
PSLILWQ"
complement(13877. .14227)
/gene="all2026"
complement(13877. .14227)
/gene="all2026"
/note="ORF_ID:all2026"
similar to nitrile hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73725.1"
/db_xref="GI:17131117"
/translation="MSEQTKRKELEAQIIVQAWKDETYRQELLNNSKAVIPEEFAIO
LPEINHVHVEENDSNFIVIPARPNEDELSEEQLEAVAGSGISDGIFFTVEKYIYDR
LKDQGEIYKAIVD"
complement(14385. .14720)
/gene="all2027"
complement(14385. .14720)
/note="ORF_ID:all2027"
similar to nitrile hydratase"
/codon_start=1
/transl_table=11
/protein_id="BAB73726.1"
/db_xref="GI:17131118"
/translation="MSEQTKRKQVDAQIIVQAWKDEAYRQELLNPNPKKIVQEQFGVQ
LPEGITVHVHVEENASNLVIFVIPARPNEDELSEEQLEAVAGGALLWTLTLLLTPIAHG
ALEHNSRK"
complement(14854. .15042)

gene

```

CDS
/gene="asl2028"
complement(14854..15042)
gene="asl2028"
/note="ORF_ID:asl2028
unknown protein"
codon_start=1
/transl_table=11
/protein_id="BAB73727.1"

Alignment Scores:
Pred. No.:      3.12e-07      Length:      334520
Score:          283.00       Matches:    154
Percent Similarity: 35.94%   Conservative: 85
Best Local Similarity: 23.16% Mismatches: 235
Query Match:      8.92%     Indels:     192
DB:               1        Gaps:       25

US-09-857-669-2 (1-609) x AP003588 (1-334520)

QY 81 GluValLeuAspLysGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
||||| :||::||| ||| |||
Db 318979 GAAGTCCTCGTAAGACCGCAATCAGGACAACATACTCCGGAAATTAGAAAACCAAGTTTAT 319038
||||| :||::||| ||| |||

QY 101 ThrMetLeuArgSerLys::----- 106
||||| :||::||| ||| |||

Db 319039 AAGGTANTTCGTACTCACGCCAGGACGGCAACACCACCGTTCCTCCAGTTACAAGAAGATATT 319098
||||| :||::||| ||| |||

QY 107 -----GlyTyPheSerSerLysValSerLeuThrGluLys----- 118
||||| :||::||| ||| |||
Db 319099 AACGCATTCTTTGGCACAGGCTTTTTCTCAA-CGTCCAAGCATCACCAGAAGATACACC 319157
||||| :||::||| ||| |||

QY 119 ---AspGlyAlaTyThrValHisIleThrProGlyProAtqThrLysIleAlaAsnVal 137
||||| :||::||| ||| |||
Db 319158 ATTAGGGTGCGAGTCACCTTCATCGTCACGCCAACCCCGCTCTTAACCAAGT----- 319211
||||| :||::||| ||| |||

QY 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyThrArgAsn 157
||||| :||::||| ||| |||
Db 319212 ----GlnProValGlyLys-----Asp 169
||||| :||::||| ||| |||

QY 158 AlaLeuGluAsnTrpGln-----GlnProValGlyLys----- 177
||||| :||::||| ||| |||
Db 319218 TCAAGCAACCTCGGACTAAGCTTCCTCAGTCCTACCCCAGGCTACTGCTGATGAAT 319277
||||| :||::||| ||| |||

QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSer----- 181
||||| :||::||| ||| |||
Db 319278 TTTCGCGCTCAGTAGTGCAGAAATTTCAACTTCGCGGATTTTACAAGAAGGGATTAAAGGA 319337
||||| :||::||| ||| |||

QY 182 ValLeuGlyAlaValThrArg-LysGlyTyThrProLeuAlaLys---LeuGlyAsnThrAr 200
||||| :||::||| ||| |||
Db 319338 ATTAACCAACGTTATCAAGACCAAGCTTAGGTTCTCGCAATGTTGTAGCGCGCCCCA 319397
||||| :||::||| ||| |||

QY 200 gAlaAlaValAsnProAspThrAlaThrValaspLeuAsnVal----- 214
||||| :||::||| ||| |||
Db 319398 GGTTCGCAAAAT-----GGAGTTGTCACCTCGCAAGTAGCCGAAGGGGTCGTTGA 319448
||||| :||::||| ||| |||

QY 215 -----ValValaspSerGlyArgPr 221
||||| :||::||| ||| |||
Db 319449 GAATATTACGTCGCTTTCCGACAAAGAAGACAGGATGTCAACCGCAACACCGTCCAGAACCAACC 319508
||||| :||::||| ||| |||

QY 221 oileAlaPheGlyaspPheGluIleThrGlyThrGlnArgTyThrProGluGlnIleValse 241
||||| :||::||| ||| |||
Db 319509 AAAT-----CGGGACGACACAGGAGTATATCATCATC 319541
||||| :||::||| ||| |||

QY 241 rGlyLeuAlaArgPheGlnProGlyThrProTyAspLeuAspLeuLeuLeuAspPheGl 261
||||| :||::||| ||| |||
Db 319542 CGCAGAAGTGGAATTTGAACACAGGACAGGTATTCAACCGCAACACCGTCCAGAACAGACCT 319601
||||| :||::||| ||| |||

QY 261 nGlnAlaLeuGluInAsnGlyHisTySerGlyAlaSerValGlnAlaAspPheAspAr 281
||||| :||::||| ||| |||
Db 319602 ACAACCGGTATTCGGGACAGATGTTTGAAGAATGCAATGTTTCCTTGCAC----CCCGG 319658
||||| :||::||| ||| |||

QY 281 gLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLe 301
||||| :||::||| ||| |||

```

```

QY 586 eserPheaspIlelaTyrgly---HisseraspIysLysIleArgTtpHisIleSerLe 605
||| ||||| ::| ||| ||| ::|
Db 320739 TCGATTATGAC-----TATGTTATCAACGATGACGGTGATACCGCATCAATTCGGTAT 320792

QY 605 uGlyThrArgPhe 609
||| |||||
Db 320793 TCGAGAAAGGTTT 320805

RESULT 34
AE005835
LOCUS AE005835.1 GI:13422995
DEFINITION Caulobacter crescentus CB15 section 161 of 359 of the complete genome.
ACCESSION AE005835
VERSION AE005835.1
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 10301)
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,M.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Hait,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Emolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
12173698
11259847
REFERENCE
AUTHORS
2 (bases 1 to 10301)
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Hait,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J.,
Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
Emolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
Fraser,C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 10301
/organism="Caulobacter crescentus CB15"
/strain="CB15"
/db_xref="taxon:190650"
83..439
/gene="CC1599"
83..439
/gene="CC1599"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23578.1"
/db_xref="GI:13422996"
/translation="MVMVTSGAERRAHPMPAARKIIVDDPPRSWKASLIDVAERGG
RISAGTASPDDTFVFDVAGGRVHLANVVRWSGTGVGVQAATQRTGPRAGGAAGAL
EIARRPLATPAEDDA"
complement(538..1515)
/gene="CC1600"
complement(538..1515)
/gene="CC1600"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23579.1"

/db_xref="GI:13422997"
/translation="MLLEIRHVTVQHYERPVRSLMELMWMQIQKTARQLRVSELDLN
PAAQSYADSGNAYHFDVPOPHDNLTIIRASAVETSGEPGPDQLDGNWDRFLRS
FYGDCDFDLPHGPHVTRTEALQTFINHDLDLKRDRPLTAVRSLSRTEIYGAFTQ
PGVTADCSLDLTLGAGRCVQDFAIHMLVCRGMPVPARYVSGYLFETDADGRSDP
DATHAWVEVFLPLSRWVGFDPNTNMAGERHVAVAGRDYADVTPSRGVTKGDAESQL
AVGVSVRRAALSLRAPEFLRMAPSPFAGGRRRHDTALREELHQOQOQOQ"
complement(1527..2474)
/gene="CC1601"
complement(1527..2474)
/gene="CC1601"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23580.1"
/db_xref="GI:13422998"
/translation="MMIARVADSLYWLGRYTERAEHLISRLSSVWMLNATLQDTAGAAQ
WTMAAAGAPEDGAGVGFDAAGSHALVDRSDPNVSSVSLARARENARQVRDQTTTET
WERLNLLYKVDTPNATKDFAGSDIFLHDVIADLHLFKGAADTMSHGESRPMIG
IYMERALYSRLLEVCFAESPHTGRIGDHVALSVLRMACALEPILRVYTAETIEPRHI
LEFLVDEDFPRSIRFATAQIEQHLSAMARVSANERAGPERLAGRLKARLQFADVDE
LEAVGAGPLLTVVNECARIHEAIYETFEVAYPLETLPA"
complement(2471..4108)
/gene="CC1602"
complement(2471..4108)
/gene="CC1602"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23581.1"
/db_xref="GI:13422999"
/translation="MNRNLAUVESKQERVALESGRMPGLGQAQDRQRDRSVAAKT
QTLDATPTLPMTEAAYLPGVADMEPTGEVROHYDPLHGRMSTLGAEEALAAQRTFL
ERSFLQGITFTVYGAENTERIIPDLPFLIPASEAWRIEAGTLQRLALNLFLLAE
IYGDQDILMDGVVPRELVIGAPS YRREOHLYVPHKAYANCYGDJLRCDQDGFVLE
DNLRFVSGVSYMLANLDAARFTPGTYRAAGVVRPVRYDPLLSTLKSMAADWRADPQ
VVLTPGVTVNSAYEHAYLARLMGVPLVEGRDLYVHNNVYMTTGLRIDITVYRV
DDDFIDPLFPRDSISGAAGLFNAYRAGNVVICNAGPGVADDAKAVAYVDPDITVYV
GEDAIPLETLYLCREPROLSHLANLQKLVKVGASGSGYGLVGPASOKERDDFA
LAKADPENYIAQPTIQLSTAFCLVDGRLEPHRVDLRFILSGEKTIVTTCGALTVAL
KRGLSVNSSGGGGSKDVTWVLADAPSNNGHGSARP"
4142..6028
/gene="CC1603"
4142..6028
/gene="CC1603"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23582.1"
/db_xref="GI:13423000"
/translation="MRSARDGGGRHSERDTSAPDYSRGGGASRGSRDLVLGRTTLL
AAVTWLSASAAMADEPMAOTOGVEDRALDAIORALSDSKQPPRSSEARRRRARQGE
DALAVLRAEGYAYTYVEPDVTEGDPRAIVRITPGPAFLADPHIDMSGSPDPDGVVQ
RAVAAMRLTEGEPGRSADVAGRGRIVAQVAKGLYADVAAPREVVDHADRTVTRPF
RINAGELVRLNGVDVYTKRTNPENWVGRLAHPVAGDVYDPEDVAELERLDRDAYIDS
ISVSLAGTDRASAEGRPVVVTLSDRRARTIELGAGYSTSEGAGVDARWIRYNNOKRA
DTTYALRFKLEQRLGABISLPHRRRPOQTLLKNSVSFRNDTDAYNETGATGVGDLT
RRKOTYAFRTFGVSDLSQTEQVNRNGLIAGRLKLNATLAGLAAYMAYSDSDILDPK
RGWRLETRATPTVAGDTVPYKLAGQSGAYLPFGKQDSTVLAARVAKLAIGLAGLL
FDPASRRFSGGGGSVRGVAYQAIGPRLSDNTPOGCSILVETSEVPOKLTDRWSGVA
FVDAAGLGTHTTQREDFRAGAGLVGVYDGLGFPDIRADIAAPLGRRKGDPKFOIYLSI
GOSP"
6025..10221
/gene="CC1604"
6025..10221
/gene="CC1604"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```



```

Db 5798 GAGGTCCGCCAGAGATCACCGATCGCTGGAGCGCGCTCGCTTCGATCGCGCGCT 5857
Qy 560 AlaAlaAlaAsnPhe-----LysArgMetLeuLeuLysHisClySerClyLeuGlyVal 577
Db 5858 ATCGGAGCCCATGAACCCCGAGCGAGGACTTCGCGCTGGCGCGCGCTTCGCGCTC 5917
Qy 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAla-----TyrGly 593
Db 5918 CTTACGATCTGGGCTTTGGACCGATCCGGCGCGACATCGCGCGCTGGAGCGCGC 5977
Qy 594 HisSerAspLysLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 5978 AAGGGCGAGCCGAGAGTCCAGATCTATCTCAGCATCGGCGAAGCTTT 6025

RESULT 35
LOCUS NGU81959 2379 bp DNA linear BCT 24-JAN-2002
DEFINITION Neisseria gonorrhoeae outer membrane protein (omp85) gene, complete
cds
ACCESSION U81959
VERSION U81959.1 GI:1766041
KEYWORDS
SOURCE
ORGANISM
.
Neisseria gonorrhoeae.
Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
1 (bases 1 to 2379)
REFERENCE
AUTHORS Manning,D.S., Reschke,D.K. and Judd,R.C.
TITLE omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis
are similar to Haemophilus influenzae D-15-Ag and Pasteurella
multocida Oms87
JOURNAL Microb. Pathog. 25 (1), 11-21 (1998)
MEDLINE 98379445
PUBMED 9705245
REFERENCE
AUTHORS Reschke,D.K., Manning,D.S. and Judd,R.C.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1996) Division of Biological Sciences, University
of Montana, Health Sciences 104, Missoula, MT 59812-1002, USA
FEATURES
source
1. .2379
/organism="Neisseria gonorrhoeae"
/strain="FA19"
/db_xref="taxon:485"
1. .2379
/gene="omp85"
1. .2379
/feature="omp85"
/feature="omp85"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AAC17600.1"
/db_xref="GI:1766042"
/transl_table=11
/translation="MKLQIASALMLGLISPLAFADFTIQDIRVEGLQTEPSTFVNY
LPVKGVDYNDTHGSAIKSLVATGTFDDVRVETADGLLLTVICPTIGSLNITGAK
MLQNDLKKLSEFSGLAQSOYFNOATLNOAVAGLKEEYLGRLKLNLIQTPKATKLARN
RVDIDITDEKSAKLTIDIEFGNNOVYSDPKLWMSITTEGGIWTMLTRSDRFDROK
ADOMKEVTPYONKQYDFEILDTIDOTNEDKTROTITITVHEGGRFWKYSLEGDT
NEVPKALEKLLTNKFKWYEQONTAVLGEONRMGSAGYATSEISVQPLPNKATKT
VDFVLHIEPQRKIYVNEIHITGNKTRDEVRELRQMSAPYDTSKLQSKSERVELL
GYFQNVQFDPLAGTPDKVDLNMSTERSGSLDSAGWVODTGLVMSAGVSDONLF
GTGASLARASRSKTTLNGSLFTDPTADGVSIGYDIYKAFDPRKASTSVKQYKT
TTAGGVRMGIPVTFDRVNFGLAAEHLTVNTYKAPRYADFIKQYKTFDGDGSPK
GLLYKGTVMGRNKTDSALWPTRGYLTGVNAELALPGSKLOYYSATINOTWFWPLSKT
FTLMLGEGVIGAGYGRTEKIIPFENFYGGISVGRVYEGCTIGPKYVDEWFEKISVG
GNKKANYSALLPMPGAKDARTVRLSLPADAGSVWDGRTTYTAENGNNKSVISENAH
KSTFTNLRISAGAVTWLSPGPNKFIYAYPLKKKPEDEIQRFQOLGTTTF"
BASE COUNT 636 a 752 c 592 g 399 t
ORIGIN
Alignment Scores: 1.55e-08 Length: 2379
Pred. No.:

```

```

Score: 260.50 Matches: 145
Percent Similarity: 36.55% Conservative: 86
Best Local Similarity: 22.94% Mismatches: 245
Query Match: 8.21% Indels: 157
DB: 1 Gaps: 28

US-09-857-669-2 (1-609) x NGU81959 (1-2379)

Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys-----ThrMet 102
Db 626 GAAGGACCGGTTCGACCGCCAGAAATTCGCCCAAGACATCGGAAAGTAACCGACTTCT 685
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 686 ACCAGAACAAAC--GGCTACTTCGATTCCTCGTATCTCGATACCGACATCCAAACCAACGAA 744
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAACACGAGCAGCATCAAAATCACCCTCCACGAAGCGGAGCTTTCGCTGGGCG 804
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyr 155
Db 805 AAAGTGTGCAAT-----GAGGCGACACCAACGAAAGTCCCC 840
Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnGlnProValGly 167
Db 841 AAGGCGGAACTCGAAAACACTCTCTGACCATGAAGCCCGCAATGGTACGAA----- 891
Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCGCTTTTGGTGAGATTACG 927
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 928 AACCGCATGGGCTCGGCGAGCTACGCATACAGCGAAATC---AGCTACAGCGCGTGGCG 984
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAla 223
Db 985 AAGCGCGGAAACAAACCGTCGATTCCTCTGCACATCGAACCGGCGAGAAAATCTAC 1044
Qy 224 PheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1045 GTCACGAAATCCATCATCCGCGCAACAAACACCCGCGAGAGTGTGGCGCGCGAA 1104
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
Db 1105 TTGCGCCAAATGGAATCCGCGCTTACGACACCTCCAAAGCTGCAACGCTCCAAAGAGCG 1164
Qy 264 LeuGlnGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1165 GTCGAGCTTTTGGGCTACTTCGAC-----AACGTACAGTTTGTATCGCTCCGCTGGC 1218
Qy 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACGCGCGCAAAAGTCGATTTAAATGAGCGCTGACCGAGCTTCCACCGGCTCGCTC 1278
Qy 302 GluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
Db 1279 GACTTGAGCGCGCGGTGGGTTCAGTACCGCTTGGTTCATGTCCCGCGCGGTATCGCAG 1338
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
Db 1339 GACAACCTTTTCGGTACGGGC-----AAGTCGCGCGCGCTTGGCGCGCTCGCAAGCAA 1392
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 1393 ACCACCTCAACGGCTCGCTGCG-----TTTACCACCGCGTACTTTCACCGCA 1440
Qy 362 -----AsnValSerTyrAsn-----AfrSer 368
Db 1441 GACGGGGTTCAGCTGGCTACGATATTTACGAAAGGCTTCGACCGCGCAAGCATCG 1500
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381

```

```

Db 1501 ACCAGCGTCAACAAATATAAACCACCGCGGGCGGCTAAGGATGGGTATCCCC 1560
Qy 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGCAATACAGCGCGTCAATTTTCGGGCTGGCGGGAACACCTGACCGTCAACACC 1620
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgGlyIleProGly 408
Db 1621 TACAACAAGACCCCAACAGCTATCCCGACTTATCAACAACATACGCAAAACCGAGCG 1680
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1681 GCAGAC-----GGCAGCTTCAAGAGCGCTGTGTACAAAGCAGCATCTCGCTGCTG 1728
Qy 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GGGCGCAACAAGACGACAGCGCTTATGGCGGACGCGCGC---TACCTGACCGCG--- 1782
Qy 446 IleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAAATCGCGAAATCGCGCTCGCGCGCAGCAAA 1815
Qy 466 AlaGlyTyrPhePheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAATACTACTCTCGCGCACCCACCAACCACTGGTCTTCCTCCCTTAACCAAACTTC 1875
Qy 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACCTGATCTCTCGCGGGAAGTCGCGCATTCGCGGGCGGTACGGCAGAACCAAGAAATC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCCTCTTTGAAACTTCTACGCGCGCGCTGGTTCGGTTCGGCGGTACGAAACCGCG 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACCTCGCGCGGAAAGTGTGACGAATACGCGGCAAAATACGTACGCGCGCAAC--- 2052
Qy 527 SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCCAAACGCTCTCCGCGAGCTGTCTTCGCGGTACGCGGTGCG 2100
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACGACGCGCTCGCGCTGAGCGTGTTCGCCGCGGACGACGCGTGTGGGACGCG 2160
Qy 560 -----AlaAlaAlaAsnPheLysArgMet----- 567
Db 2161 AGAACCTATACCGCGCGGCAAAACGCTAACCAAAATCGGTTTACTCGGAAAACGCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2221 AATCCACCTTTACCAACCAATTTGCGCTATTTCGCCGCGGCGCGGTACTCGCTCG 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CCTTTGGCGCGATGAATTCATCTACGCTTACCGCTCGGTAAGAAAAAACGGAAGACGAA 2340
Qy 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCCAAACGCTTTCATTCAGCTCGGACGACGCTTC 2376

```

```

RESULT 36
AX155443
LOCUS 2379 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 6 from Patent WO0138350.
ACCESSION AX155443
VERSION AX155443.1 - GI:14536771
KEYWORDS
SOURCE
ORGANISM
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;

```

```

Neisseria.
1 (bases 1 to 2379)
AUTHORS Giuliani, M., Pizzo, M., Rappuoli, R. and Holst, J.
TITLE 85kda neisserial antigen
PATENT: WO 0138350-A 6 31-MAY-2001;
Chiron Spa (IT); Statens Institut for Folkehelise (NO)
FEATURES
Location/Qualifiers
source 1..2379
Organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 636 a 755 c 596 g 392 t
ORIGIN
Alignment Scores:
Pred. No.: 2,87e-08 Length: 2379
Score: 256.50 Matches: 144
Percent Similarity: 36.39% Conservative: 86
Best Local Similarity: 22.78% Mismatches: 246
Query Match: 8.08% Indels: 157
DB: Gaps: 28
US-09-857-669-2 (1-609) x AX155443 (1-2379)
Qy 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys-----ThrMet 102
Db 526 GAAGCCAGCGGTTCGACCGCCAGAAATTCGCCCAACATCGCCCAAGACTTACCGACTTCT 585
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 586 ACCAGAACAC-CCGTACTTCGATTTTCGTATTCGTATTCGATACCGACATCCCAACACGAA 744
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAACACGCGACGACCATCAAAATACCGTCCACGAGCGGACGCTTTCGCGTGGGC 804
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAAGTGTCTGATT-----GAAGGCGACACCAAGTCCCG 840
Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnProValGly 167
Db 841 AAGCGCAACTCGAAAACTGCTGACCATGAGCCCGCAATGCTACGAA----- 891
Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCCAGCAGATGACCGCGCTTTTGGGTGAGATTCAG 927
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 928 AACCGCATGGGCTCGCGAGCTACGCATACAGCGAAATC---ACGCTACAGCGCGTCCG 984
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223
Db 985 AACCGCGGAACCAAAACCGTCTGATTTCTGTCGACATCGAACCGCGCGGCAAAATCTAC 1044
Qy 224 PheGlyAspPheIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1045 GTCAACGAATCCACATACCGCGCAACAAACCCCGCGAGTCTGTCGCCCGCGAA 1104
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
Db 1105 TTCGCCAAATGGAATCGCGCTTACGACACCTCCCAAGCTGCAACGCTCCAAAGACGCG 1164
Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1165 GTCGAGCTTTTGGGTACTTCGAC-----AACGTACAGTTTTCGATCGCGCTTCGCG 1218
Qy 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACGCGCGCAAAAGTCGATTTGAACATGACCTGACCGACCGCTCCACCGGCTCGCTC 1278
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321

```

Db 1279 GACTTGTACGGCGGCTGGTTTCAGGTATACCGGCTTGGTTCATGTCCGCGCGCTATCCGAG 1338
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValThrAspMetAspLysTyrGlu 341
|||||
Db 1339 GACAACTGTTCGGTACGGC-----AAGTCGGCGCCCTCGCGCCCTCGCGAAGCAAA 1392
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
|||||
Db 1393 ACCAGCTCAACGGCTCGCTGTCG-----TTTACCGACCGCTACTTTCACGGCA 1440
Qy 362 -----AsnValSerTyrAsn----- 381
Db 1441 GACGGGTACCGCTGGCTAGCATATTACGGAAAGCCTTCGACCCCGCAAGCATCG 1500
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
Db 1501 ACCAGCGTCAACAATATAAACACCCACCGCGCGCGGTAAGGATGGGTATCCCC 1560
Qy 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGAATACGACCGCTCAATTTCGGGCTGGCGCGGACACCTGACCGTCAACACC 1620
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGly 408
Db 1621 TACAACAAAGCACCAACGCTATCGCGACTTTATCAGAAATACGGCAAAACCGACGGC 1680
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1681 GCAGAC-----GGCAGCTTCAAGCGCTGCTGTACAAAGGACCGCTCGCGTGG 1728
Qy 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrIleAspGlyLys 445
Db 1729 GGGCCCAACAGACGACGCGCTATGCGCGCGCGC-----TACCTGACCGCGC--- 1782
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAATGCGCGAATCGCCCTGCCCGCGCAAAA 1815
Qy 466 AlaGlyTyrPheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAACTACTCTCGGCACCCACCAACCTGCTTCTCCCTTAAGCAAAACCTTC 1875
Qy 479 ThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1876 ACCTGTATGCTCGCGCGAAGTGGCGGCTTTCGGCGGCTAGCGGCAACCAAGAAATC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCTCTTTGAAACTTCTACGGCGCGCTGGTTCGGTTCGGTTCGCGGCTACGAAAGCGC 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACCTTCGGCGGCAAGTGTATGAGATACGGCGAATAAATACGCTACGGCGGCAAC--- 2052
Qy 527 SerValProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCAACGCTCTCCGCGAGCTGCTCTCCCGATGCTCCGCTCG 2100
Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAAGACGACGACCGCTCGCTGAGCTGTGTTCGCGACGACGCGCTGTGGGACGCGC 2160
Qy 560 -----AlaAlaAlaAsnPheLysArgMet----- 567
Db 2161 AGAACCCTATACCGCGCGGAAACCGTAAACAACAAATCGGTTTACTCGGAAACCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2221 AAATCCACTTTTACCAAGCAATTCGGCTATTCGCGCGCGCGCTTACCTGGCTCTCG 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CCTTTGGTCCGATTAATTCAGCTACGCTACCCCTGAGAGAAAAACCGGACGCA 2340

Qy 598 LysIleArgTyrPheHisIleSerLeuGlyThrArgPhe 609
|||||
Db 2341 ATCCAACGCTTCCAATTCACGCTCGGCACGACGCTTC 2376
RESULT 37
LOCUS AX202493 2379 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 6 from Patent WO0152885.
ACCESSION AX202493
VERSION AX202493.1 GI:15392216
KEYWORDS
SOURCE Neisseria gonorrhoeae.
ORGANISM Neisseria gonorrhoeae.
REFERENCE 1. (bases 1 to 2379)
AUTHORS Pizsa M.C., Rappuoli, R.C. and Giuliani, M.C.
TITLE Outer membrane vesicle (omv) vaccine comprising N. meningitidis
serogroup b outer membrane proteins
JOURNAL Patent: WO 0152885-A 6 26-JUL-2001;
Chiron Spa (IT)
FEATURES
Location/Qualifiers
source 1. 2379
/organism="Neisseria gonorrhoeae"
/db_xref="taxon:485"
BASE COUNT 636 a 755 c 596 g 392 t
ORIGIN
Alignment Scores:
Pred. No.: 2,87e-08 Length: 2379
Score: 256.50 Matches: 144
Percent Similarity: 36.39% Conservative: 86
Best local Similarity: 22.78% Mismatches: 246
Query Match: 8.08% Indels: 157
DB: 6 Gaps: 28
US-09-857-669-2 (1-609) x AX202493 (1-2379)
Qy 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys-----ThrMet 102
Db 626 GAAGGACCGGTTCCGACCGCGCAAAATTCGCCAAGACATCGAAAGTAACGACTTCT 685
Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
Db 686 ACCAGAACAC--GGCTACTTCCGATTCCTCGATACCGACATCCCAACCAACGAA 744
Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
Db 745 GACAAACACCGCACCATCAAAATCAGCTCCACCAAGCGGAGCTTCGCTGGGCGC 804
Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
Db 805 AAAGTGTGCTAT-----GAAGGCGACACCAACGAAAGTCCCC 840
Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnGlnProValGly 167
Db 841 AAGCGCGAACTGGAAAACTGCTGACCATGAGCCCGCAATGGTACGAA----- 891
Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
Db 892 -----CGCAGCAGATGACCGCGCTTTTGGGTGAGATTCTAG 927
Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
Db 928 AACCGCATCGGCTCGGAGGCTAGCATACAGCGAAATC---ACGCTAGACGCGTCCGC 984
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAla 223
Db 985 AACCGCGAAACCAACCGCTGATTCGCTGACATCCGACCGCGGCGGAAATCTAC 1044
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
|||||

Db 1045 GTCAACGAAATCCACATCACCGGCAACAAACAAACCGCGACGAAGTCTGTCGCGCGGAA 1104
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAla 263
Db 1105 TTCGGCAAAATCGAATCCGGCGCTTACGACACCTCCAAAGTCCAAAGCGGC 1164
Qy 264 LeuGlnGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1165 GTGAGCTTTTGGGCTACTTCGAC-----AACGTACAGTTTGTATGCCCTCCGCTTGC 1218
Qy 284 Gly-----AspArgValProValValSerValThrGluValLysArgHisLysLeu 301
Db 1219 GGTACCGCGCAAGTGAATTTGAACATGAGCTGACCAACGCTCCACCGCGCTCGCTC 1278
Qy 302 GlyThrGlyLeuArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
Db 1279 GACTTGGCGCGCGGTTCAGATACCGCTTGTGTATGTCGCGCGGTATCGCAG 1338
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
Db 1339 GACAACCTGTCGGTACGGG-----AAGTCGCGCGCGCTCGCGCGCTCGCGAAGCAA 1392
Qy 342 ThrThrLeuAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
Db 1393 ACCAGCTCAACGGCTCGCTGTCG-----TTTACCGACCGGTACTTCAGCGCA 1440
Qy 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1441 GACGGGTGCTCGCTGGTACGATATTATCGGAAGACCTTCGACCGCGCAAGCATCG 1500
Qy 369 ThrThrGlnAsnLeuGlyLysArgAlaPheSerGlyGly----- 381
Db 1501 ACCAGCTCAACAAATATAAACACCAACCGCGCGCGCTGAAGATGGTATCCCG 1560
Qy 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1561 GTTACCGAATACAGCGCGTCAATTCGGCTGCGCGGGAACACCTGACCGTCAACACG 1620
Qy 394 -----ArgLeuGlyAlaGluPheLeuAlaGlyArgLysIleProGly 408
Db 1621 TACAACAAAGCAACCAACCGTATCGGACTTTATCAGGAATACGGCAAAACCGACGC 1680
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1681 CGACG-----GGCAGCTTCAAGCGCTGTGTACAAAGCACCGCTCGCGCTGG 1728
Qy 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1729 GCGCGCAACAGACGACGCGCTCATGCGCGACGCGCGC---TACCTGACCGCGC--- 1782
Qy 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1783 -----GTAAATGCCAAATCCCTCCCTCGCGCGCGCAAA 1815
Qy 466 AlaGlyTyrPhePheThrProGluAsnLys-----LysLeuGly 478
Db 1816 CTGCAATACTACTCGCGCACCCACACCAACCTGGTTCCTTCCTTAAAGCAAAACCTTC 1875
Qy 479 ThrPheIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAla---AspVal 497
Db 1876 ACCTGATGCTCGCGCGGAAGTCCGATTCGGCGGCTACCGCGAGACCAACCAAGAAATC 1935
Qy 498 ProSerGlyLeuMetPheArgSerGlyAlaSerValArgGlyTyrGluLeuAsp 517
Db 1936 CCGTCTCTTTGAAAACTTCTACGGGGGGGCTTGGTTCGGTTCGGCGCTACGAAAGCGCG 1995
Qy 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 1996 ACCTCGCGCGGAAGTGTATGACGAATACGGCGCAAAATACGTACCGCGCGCAAC--- 2052
Qy 527 SerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2053 -----AAAAAGCCAAAGTCTCCGCGCGAGTGTCTTCCCGATGCCCGGTGCG 2100

Qy 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2101 AAGACGCGACGCGCTCGCGCTAGCGCTGTTCGCCAGCGACGCGGTGTGGGACGCG 2160
Qy 560 -----AlaAlaAlaAsnPheLysArgMet----- 567
Db 2161 AGAACCTATACCGCGCGCGCAAAACGGTAACAACAATCGGTTTACTCGGAAACGCGCAT 2220
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2221 AATCCACCTTTTACCAACGAATTCGGCTATTTCGCGCGCGCGGTTCACGTGCTCGC 2280
Qy 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----GlyHisSerAspLys 597
Db 2281 CCTTTGGTCCGATCAATTCAGCTACCGCTACCGCTGAAGAAAAACCGCAACGCA 2340
Qy 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2341 ATCCAACGCTTCCAATTCAGCTCGGCACGAGTTC 2376
RESULT 38
AX155447
LOCUS 2394 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 10 from Patent WO0138350.
ACCESSION AX155447
VERSION AX155447.1 GI:14536772
KEYWORDS
SOURCE Neisseria gonorrhoeae.
ORGANISM Neisseria gonorrhoeae.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Giulianini, M.M., Pizzi, M., Rappuoli, R. and Holst, J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 10 31-MAY-2001;
Chiron spa (IT) ; Statens Institut for Folkehelse (NO)
FEATURES
Location/Qualifiers
source 1..2394
/organism="Neisseria gonorrhoeae"
/db_xref="taxon:485"
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
Alignment Scores:
Pred. No.: 4.61e-08 Length: 2394
Score: 253.50 Matches: 151
Percent Similarity: 36.50% Conservative: 99
Best Local Similarity: 22.04% Mismatches: 276
Query Match: 7.95% Indels: 159
DB: 6 Gaps: 27
US-09-857-669-2 (1-609) x AX155447 (1-2394)
Qy 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
Db 470 TCGCGCGCAACCGCGCTCGACATCGACATCGACGAGGCAAAATCCGCCAAATCA 529
Qy 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
Db 530 CCGACATCGAATTTGAAGCAACCAAGTCTATTCCGACCGCAAACTGTGCGGCGAGATGT 589
Qy 73 oLeuLethrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaG1 93
Db 590 CGCTGACCGCAAGCGCGCATTTGGACATGGCTGACACGACCAACCAATTCACGACGAGA 649
Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe- 109
Db 650 AATTTGCCCGCAGACGATGGAAGAAAGTAACCGGACTTCTACCAAGAACGCGCTACTTCGATT 709
Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis1 126
Db 710 TCGCGATCTCGATACCGACATCCAAACCAACGAAGCAAAACCAACGACGACCATCAAA 769

QY 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
 Db 770 TCACCGCTCCAGAAAGGGAGCTTTCGGTGGGGCAAGCTCCCATC----- 816
 QY 144 spIleuSerAspGlyAsnLeuAlaGlyTyrTrpArgAsnAlaLeuGlu----- 160
 Db 817 -----GAAGGGACACCAACAGTCCCAACAGCCCAAGCTGGAAGAACTGCTGA 865
 QY 161 -----AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpG 176
 Db 866 CCATGAAGCCCGCAAAATGTACGA-----C 892
 QY 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
 Db 893 GCCAGCAGATCACCGCGCTTTGGTGAGATTCCAGAACCGCATGGCTCGCGAGGCTACG 952
 QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
 Db 953 CATACGCGGAATC---AGCTACAGCCGCTGCCAACCGCGAAACCAAAACCGTCGATT 1009
 QY 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
 Db 1010 TCGTCTGCATCATCAACCGCGCGGAAATCTACCTCAACGAAATCCACATCACCGGCA 1059
 QY 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
 Db 1070 ACAACAAACCGCGAGCAAGTCTGCGCGCGCGAAATGCGCCAAATGGAATCGCGCGCTT 1129
 QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
 Db 1130 AGCAGACCTCCAAAGCTCCAAAGAGCGCGCTCGAGCTTTTGGCTACTTCGAC- 1188
 QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
 Db 1189 -----AAGTACAGTTGATGCGCTCGCGCGCACCCGCAAAAGTCGATTGGA 1243
 QY 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
 Db 1244 ACATGAGCTGACCAAGCTTCCACCGCGCTCGCTCGACTGAGCGCGGCTGGGTACAGG 1303
 QY 310 luTyrGlyLeuGlyGlyLysIleAlaLysTyrTrpAsnLeuPheAsnLysGlyTyrI 330
 Db 1304 ATACCGGCTGTGTATGCTCCGAGGCGCTTCCCAAGACACCTGTTTCGGTACGGGC---- 1359
 QY 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
 Db 1360 --AAGTCGCGCGCTGCGCGCTCCAGAAAGCAAAACACGCTCAACGGCTCGCTGCT 1417
 QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
 Db 1418 TTACCGACCGTACTTCACGACAGCGGGTTCAGCTCGGCTGACGATGTTTACGAAAG 1477
 QY 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
 Db 1478 CTTTCGACCGCGCAAGACATCGACAGCATCAACAATATAAAACCAACCGCGAGCG 1537
 QY 373 euGluLysArgAlaPheSerGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
 Db 1538 CAGGCATCGCATGAGCTGCTTTTACCGAATACGACCGCGGAATTCGGTGGTGGG 1597
 QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
 Db 1598 CAGAACCTGACCGCTCAACACCTACAACAAAGCGCCCAACATCTATGCGGCTTATCA 1657
 QY 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
 Db 1658 AGAAATACGGCAAAACCGCGGACAGAC-----GCGACTTCAAAAGGCTGGCTGT 1708
 QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
 Db 1709 ACAAAAGGTACCGTGGCTGGGGCGCAACAAACCGACAGCGGCTTATGCGCCGACGCGG 1768

QY 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
 Db 1769 GC---TACTCGACGGC-----GTGAACGCGCGAAA 1795
 QY 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
 Db 1796 TCGCCCTCGCGCGAGCAAACTGCAATACTACTCCGCCACCCACCAACCAACCTGTTCT 1855
 QY 476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
 Db 1856 TCCTCTTAAGCAAAACCTTCACGCTGATCTCGCGCGGAAAGTCGCATTCGGCGGGCT 1915
 QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerV 511
 Db 1916 ACGGCAGAACCAAAAGAAATCCCTTCTTGAATACTTCTACGCGCGCGCTGGGTTCGG 1975
 QY 511 alaArgGlyTyrGluLeuAspSerIleGly----- 520
 Db 1976 TCGCGGATAGCAAGCGGACGCTCGGTCCGAAAGTGTATGACGANTACGGCGGAAAAA 2035
 QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
 Db 2036 TCAGCTACGGCGCAAC-----AAAAAGCAACGCTCTCCGCCGAGCTGC 2080
 QY 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
 Db 2081 TCTTCCCATGCCCGCGCGGAAAGACGCGCGACCTGCGCCTGAGCCTGTTTCCGACG 2140
 QY 557 etGlyAsp----- 559
 Db 2141 CAGCGAGCGTGTGGGACGGCAAAACCTTACGACGACAACAGCAGTTCCTCGCGACCGGCA 2200
 QY 560 -----AlaAlaAlaAsnPhelysArgMet-----LysLeuL 570
 Db 2201 GGGTTCAAAACATTTACGGCGCGGCAATACCATTAATCCACTTTTACCAACGAATTCG 2260
 QY 570 yHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
 Db 2261 GCTATTCGCGCGCGCGCTTACTTCTGCTGCTGCTTTAGCCCGATGAATTCAGCT 2320
 QY 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
 Db 2321 ACGCTACCGCTGAGAAAAAACCGGAGAGCAAAATCCAAACGCTTCCAAATTCGAAC 2380
 QY 606 lyThrArgPhe 609
 Db 2381 GCACGAGCTTC 2391
 RESULT 39
 AX202497
 LOCUS AX202497 2394 bp DNA linear PAT 30-AUG-2001
 DEFINITION Sequence 10 from Patent WO0152885.
 ACCESSION AX202497
 VERSION AX202497.1 GI:15392217
 KEYWORDS
 SOURCE Neisseria meningitidis
 ORGANISM Neisseria meningitidis
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 REFERENCE 1 (bases 1 to 2394)
 AUTHORS Pizza,M.C., Rappuoli,R.C. and Giuliani,M.C.
 TITLE Outer membrane vesicle (omv) vaccine comprising N. meningitidis
 serogroup b outer membrane proteins
 JOURNAL Patent: WO 0152885-A 10 26-JUL-2001;
 Chiron Spa (IT)
 FEATURES
 source Location/Qualifiers
 1..2394
 /organism="Neisseria meningitidis"
 /db_xref="taxon:487"
 BASE COUNT 655 a 758 c 395 t
 ORIGIN
 Alignment Scores:

Pred. No.: 4, 61e-08 Length: 2394
Score: 253.50 Matches: 151
Percent Similarity: 36.50% Conservative: 99
Best Local Similarity: 22.04% Mismatches: 276
Query Match: 7.99% Indels: 159
DB: 6 Gaps: 27

US-09-857-669-2 (1-609) x AX202497 (1-2394)

Qy 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
Dbb 470 TCCCGCGCAACCGCTGCACATCGACATGTCAGAGGGCAATACCGCCAAATCA 529
Qy 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
Dbb 530 CCGACATCGAATTTGAAGCAACCAAGTCTATTCCGACCGCAACATGATCGCGCATCT 589
Qy 73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAgl 93
Dbb 590 CGGTGACCGAAGCGGCATTTGGACATGGCTGACACGACGAACCAATTCACAGCACA 649
Qy 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyrPhe---- 109
Dbb 650 AATTTCCCAAGACATGGAAGAAAGTAACCGACTTCTACAGAACACACGGCTACTTCGATT 709
Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisI 126
Dbb 710 TCCGCATCTCCATCCGACATCCCAACCAACGAAGACAAACCAACGACGACCATCAAA 769
Qy 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
Dbb 770 TCACCGTCCACGAAGCGGACGTTCCCTGGGGCAAGTCTCCATC----- 816
Qy 144 spileLeuSerAspGlyAsnLeuAglGluTyrTyrArgAsnAlaLeuGlu----- 160
Dbb 817 -----GAAGCGACACCAACGAAGTCCCAAGACCGCAACTCGAANAACCTGCTGA 865
Qy 161 -----AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpG 176
Dbb 866 CCATGAAGCGCGCAATGGTACGAA-----C 892
Qy 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
Dbb 893 GCACGACATGACCGCGCTTTGGGTGAGATTCAGACCCGATCGGCTCGGACGGCTACG 952
Qy 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
Dbb 953 CATACAGCAAAATC---AGCGTACAGCGCTGCCAAGCGCGCAACCAACCAACCGTCGATT 1009
Qy 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
Dbb 1010 TCGTCTGCACATCGACCGCGCGGAAATCTACGTCAACGAAATCCACATCCCGCGCA 1069
Qy 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
Dbb 1070 ACAACAAAACCGCGCAAGTCTGTGCGCGCAATTCGCGCAATGGAATCGCGCGCTT 1129
Qy 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
Dbb 1130 ACACACCTCCCAAGCTGCACCGCTCCAAAGAGCGCGTTCGAGCTTTGGGCTACTTCGAC- 1188
Qy 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
Dbb 1189 -----AACGTACAGTTGATGCGGCTCCCGCTTGCAGCGACACCGCAACGATCGATTGA 1243
Qy 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
Dbb 1244 ACATGACCTGACCGAAGCTTCCACCGCTCGCTCGACTTCAGCGCGGCTGGGTACAGG 1303
Qy 310 luTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
Dbb 1304 ATACCGCGCTGTCATGTCCGCGAGCGGTTTCCCAAGACAACTGTTTCGCTACGGCG- 1359

Qy 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
Dbb 1360 --AAGTCGGCGCCCTCGCGCTCAGCAAGCAAAACACGCTCAACGGCTCGCTCGCT 1417
Qy 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
Dbb 1418 TTACCGACCGCTACTTTCACGCGACAGCGGGTTCAGCTGGCTGCTTTCGCGAAAG 1477
Qy 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
Dbb 1478 CTTTCGACCGCGCAACGATCGACGATCAAAACAAATATAAACCACACGCGAGGG 1537
Qy 373 euGluLysArgAlaPheSerGlyGlyIleTyrValArgAspArgAlaGlyIleAspA 393
Dbb 1538 CAGGCATCCGCATGAGCTGCTGTTCCGAAATACGACCGCGTGAATTCGGTGTGGTGG 1597
Qy 393 la-----ArgLeuGlyAlaGluPheLeuA 401
Dbb 1598 CAGACACCTGACCGTCAACACCTACAACAAGCGCCCAACACTATGCGGCTTTATCA 1657
Qy 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
Dbb 1658 AGAATACGCAAAACCGACGCGACAGAC-----GGCAGCTCAAAAGCTGCGTCT 1708
Qy 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnG 439
Dbb 1709 ACAAAGTACCTCGCTGGGCGCAACAACCGACGCGCTTATGCGCGACCGCGG 1768
Qy 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
Dbb 1769 GC---TACCTGACGGGC-----GTACGCGCGGAA 1795
Qy 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
Dbb 1796 TCGCCCTCGCGCGACGAACTGCACTACTCGCGCACCCACACCAACCTGGTTC 1855
Qy 476 -----LysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaA 492
Dbb 1856 TCCCTTAAAGCAAAACCTTCAGCTGATGCTCGCGCGCAAGTCGCGGCGGT 1915
Qy 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyAlaSerSerV 511
Dbb 1916 ACGGCAGAACCAAAATTCCTTCTTTGAAACTTCTACGGCGCGGCGGTTCGG 1975
Qy 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Dbb 1976 TCGCGGATACGAAAGCGCGACGCTCGGTCCGAAAGTGTATGACGAATACGGCGAA 2035
Qy 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
Dbb 2036 TCAGCTACGCGCGCAAC-----AAAAAGCAACGCTCCGCGCGAGCTCG 2080
Qy 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Dbb 2081 TCTTCCGATGCGCGCGCGCAAGACGCGCACCGCTCCGCTTTCGCGCTGTTTTCGCGAG 2140
Qy 557 etGlyAsp----- 559
Dbb 2141 CAGGACGCTGTGGCGCGCAAAACCTACGACGACACAGCAGTTCGCGCGACGGCGCA 2200
Qy 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
Dbb 2201 GGGTTCAAAACATTTACGGCGCGCGCAATACCCATAATCCACTTTTACCAAGAAATTCG 2260
Qy 570 yHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Dbb 2261 GCTATTCGCGCGCGCGGCTTACCTGGCTCTCGCTTTCGCGCTTTCGCGCTGCTTTCGCGAG 2320
Qy 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
Dbb 2321 ACCCTTACCGCTGAGAGAAAACCGGAAGACGAATTCACACGCTTCCCAATTCCAACTCG 2380
Qy 606 lyThrArgPhe 609

[illegible]

```

/EC_number="1.6.5.3"
/notes="NMA0005, nuok, NADH dehydrogenase I chain K, len:
101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase
I chain K (EC 1.6.5.3) (102 aa), fasta scores; E():
1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam
match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain K"
/protein_id="CAB83325.1"
/db_xref="GI:7378783"
/db_xref="SPTREMBL:Q9J0U9"
/translation="MITTHLVIGALGFCISAMGIFNMKNKVLVLLMSTELMLLAVN
FNFLAYSOHLGTAGQIFVEFVLTVAAESAIGLAIMVLVYENRQTIINVADLDELK"
/complement(3564..3863)
/genes="nuok"
/notes="Pfam match to entry PF00420 oxidored_q2,
NADH-ubiquinone/plastoquinone oxidoreductase chain 4L,
score 141.50, E-value 1.5e-38"
/complement(3863..4534)
/genes="nuoj"
/complement(3863..4534)
/genes="nuoj"
/EC_number="1.6.5.3"
/notes="NMA0006, nuoj, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ_RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fasta scores; E():
9.7e-21; 41.3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9JX90"
/translation="MTFOLILFYIFAVIILYGAIKTIVTAKNPVHAALHLVLTFCVSA
LMLMKAQFLGYLVLVVYGVGAVVFLFYVMNLINIEMRAGFWRPAPVAGVYVGLL
AVALIIITVNPKTDLAAAGLMDKIDPADYNNIRDLGSRIYTDLYLPFLEAAVLLGLMG
AATLALVHKRTVNPKRMDPADQVKVRADQGRMLVKMEAVKPKTESAESESVDDLKPK"
/ECGA"
/complement(3874..3879)
/genes="nuoj"
/complement(4034..4525)
/genes="nuoj"
/notes="Pfam match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"
/complement(4570..5277)
/genes="NMA0007"
/complement(4570..5277)
/genes="NMA0007"
/notes="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"

```

Alignment Scores:

Pred. No.:	3,09E-05	Length:	340806
Alignment Scores:		Matches:	151
Score:	253.50	Conservative:	99
Percent Similarity:	36.50%	Mismatches:	276
Best Local Similarity:	22.04%	Indels:	159
Query Match:	7.99%	Gaps:	27
DB:	1		

US-09-857-669-2 (1-609) x NMA122491 (1-340806)

QY 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
 ||||| ||| ||| |||||
 Db 84451 TCGCCCCCAACCGCGTCGACATCGACATTCACCATTGACGAGGCGCAATTCGCCCAAAATCA 84510

```
QY 373 euGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
DB 85519 CAGGCATCCGCGATGAGCGTCCCTGGTTACCGAATACGCGCTGAATTCGGTTTGTGG 85578
QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
DB 85579 CAGAACACCTGACCGTCAACACCTACAACAAGCGCCAAACACACTATGCCGACTTATCA 85638
QY 401 laGluGlyArgGlyIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
DB 85639 AGAATACGGCAAAACCGGCGGACAGAC-----GGCAGCTTCAAGGGTCGCTGT 85689
QY 421 eu-----ThrAlaSerTrpPlyArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
DB 85690 ACAAGGTACCGTGGCGGCGCAACAACCGACAGCGCTTATGGCCGACGCGCG 85749
QY 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
DB 85750 GC---TACCTGACGGGC-----GTGAACCCCGAAA 85776
QY 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
DB 85777 TCGCCCTGCCCGGCGAGCAAACTCAATACTACTCGCGCCACGACAAACCTGGTTCT 85836
QY 476 -----LysLeuGlyThrPheIleArgGlyGlnAlaGlyTyrThrValAlaA 492
DB 85837 TCCCTTTAAGCAAAACCTTCACCTGATCTCGGGCGGAGTGGCATTCGGGGCGGCT 85896
QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyAlaSerSerV 511
DB 85897 ACGGCAGAACCAAGAAATCCCTTCTTTGAAAACTTCTACGGCGGCGCTGGTTTCGG 85956
QY 511 alArgGlyTyrGluLeuAspSerIleCly----- 520
DB 85957 TCGCGGATAGCAAGGCGGACGCTCGGTCCGAAGTGTATGACGAATACGCGGAAAAA 86016
QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuG 540
DB 86017 TCAGCTACGGCGGCAAC-----AAAAAGCCAACTCTCCGCGGAGCTGC 86061
QY 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
DB 86062 TCTTCCCGATGCCGGCGGCAAGAGCGCGCACCTCGGCTGAGCCTGTTGCCGACG 86121
QY 557 etGlyAsp----- 559
DB 86122 CAGGCACGCTGTGGGACGGCAAACTACGACGACAACAGCAGTTCCGCGACCGCGGCA 86181
QY 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
DB 86182 GGGTTCAAAACATTACGGCGCGCGCAATACCCATAATCCACCTTTACCAACGAATGC 86241
QY 570 ysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
DB 86242 GCTATTCGGCGGCGGCGGTTTACCTGCTCTCGCCTTTAGGCCCGCATGAATTCAGCT 86301
QY 590 leAlaTyr-----GlyHisSerAspLysLysIleArgTrpHisIleSerLeuG 606
DB 86302 ACGGCTACCGCTGACAAAAAACCGGAAGACGAATCCACGCTTCCAAATCCAACTCG 86361
QY 606 lyThrArgPhe 609
DB 86362 GCACGACGTTTC 86372
```

Search completed: November 9, 2002, 02:46:21
Job time : 4583 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:03:13 ; Search time 36 Seconds
(without alignments)
2254.156 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMIKPTALLPALFFPHAY.....IAYGHSKKIRWHISLGRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 15

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002.*				
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*			
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*			
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*			
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*			
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*			
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*			
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*			
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*			
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*			
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*			
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*			
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*			
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*			
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*			
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*			
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*			
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*			
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*			
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*			
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*			
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*			
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*			
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609	100.0	609	21	AA199623
2	609	100.0	609	21	AA199624
3	249	40.9	615	21	AA199625
4	198	32.5	587	21	AA199625
5	198	32.5	615	21	AA199625
6	170	27.9	522	21	AA199625

ALIGNMENTS

RESULT 1	
ID	AA199623
XX	AA199623 standard; Protein: 609 AA.
AC	AA199623;
XX	03-JAN-2001 (first entry)
DT	
XX	Neisseria meningitidis BASR040 putative protein sequence #1.
DE	
XX	BASR040; bacterial disease; respiratory tract infection; bacteraemia; meningitis; cancer; autoimmune disease.
KW	Neisseria meningitidis.
OS	Neisseria meningitidis.
XX	WO200034430-A1.
PN	
XX	15-JUN-2000.
PD	
XX	02-DEC-1999; 99WO-EP09560.
PF	
XX	07-DEC-1998; 98GB-0026886.
PR	
XX	(SMK) SMITHKLINE BEECHAM BIOLOGICALS.
PA	
XX	Ruelle J;
PI	
XX	WPI; 2000-423426/36.
DR	N-PSDB; AAA48507.
XX	Novel BASR040 polypeptides of Neisseria meningitidis useful for diagnostic, prophylactic and therapeutic purposes against microbial diseases comprise a specific amino acid sequence
PT	
XX	Claim 4; Page 59-60; 98pp; English.
PS	
XX	The present sequence is a putative version of the Neisseria meningitidis strain ATCC13090 BASR040 protein sequence. This protein is similar to the D15 outer membrane protein of the bacterium. The protein, its gene, antibodies, antagonists and agonists can be used to diagnose and treat bacterial diseases such as those leading to upper respiratory tract infections, bacteraemia and meningitis. In addition, they can be used in vaccines for use against cancer and autoimmune diseases.
CC	
XX	Sequence 609 AA;
SQ	
Query Match 100.0%; Score 609; DB 21; Length 609;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MMIKPTALLPALFFPHAYAPAADLSENKAAGFALFKNKSPTESVKLKPFPVRIDTQ 60
Db	1 MMIKPTALLPALFFPHAYAPAADLSENKAAGFALFKNKSPTESVKLKPFPVRIDTQ 60
QY	61 DSEIKDMVEEHLPLITQQQEEVLQKEQTGFLAEAPDNVKTMLRSKGYFSSKVSLETKDG 120
Db	61 DSEIKDMVEEHLPLITQQQEEVLQKEQTGFLAEAPDNVKTMLRSKGYFSSKVSLETKDG 120
QY	121 ATVTHTPGPRTKIANVGA ILGDILSDGNLA EYR NAL ENWQOPVGSDFDDQSDWNSKT 180
Db	121 ATVTHTPGPRTKIANVGA ILGDILSDGNLA EYR NAL ENWQOPVGSDFDDQSDWNSKT 180
QY	181 SVLGAVTRKGYPLAKIGNTRAANVPDVTATDLNVVDSGRPIAFGDFEITGTORYPEQIV 240
Db	181 SVLGAVTRKGYPLAKIGNTRAANVPDVTATDLNVVDSGRPIAFGDFEITGTORYPEQIV 240
QY	241 SGLARPOGCTPYDLDLLDFQALQNGHYSGASVQADFRLQGDVFPVKVSVTEYKRRH 300
Db	241 SGLARPOGCTPYDLDLLDFQALQNGHYSGASVQADFRLQGDVFPVKVSVTEYKRRH 300
QY	301 LETGIRLDSEYGLGCKIADYDYNLNFNKG YIGSVVDMDKYETTLAAGISQPNRYRGNYWT 360
Db	301 LETGIRLDSEYGLGCKIADYDYNLNFNKG YIGSVVDMDKYETTLAAGISQPNRYRGNYWT 360

Db 301 LETGIRLDSEYGLGGKIAYDYNNLFNKGY IGSVYVMDMKYETTLLAAGISQPRNRYGNWYT 360
QY 361 SNVSYNRSTQNLKRAFGSGGIWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTQNLKRAFGSGGIWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
QY 421 LTASWKRRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKRRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
QY 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGSLE 540
Db 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGSLE 540
QY 541 YQLPFTRTLGSAGVAFHDMGDAANFKRMKLRHSGGLGYRVFSPFLAPFSFDIAYGHSDDKIR 600
Db 541 YQLPFTRTLGSAGVAFHDMGDAANFKRMKLRHSGGLGYRVFSPFLAPFSFDIAYGHSDDKIR 600
QY 601 WHISLGTFR 609
Db 601 WHISLGTFR 609
RESULT 2
AAY99624
ID AAY99624 standard; Protein; 609 AA.
AC AAY99624;
XX
DT 03-JAN-2001 (first entry)
DE Neisseria meningitidis BASB040 putative protein sequence #2.
XX
KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
KW meningitis; cancer; autoimmune disease.
XX
OS Neisseria meningitidis.
XX
PN WO200034480-A1.
XX
PD 15-JUN-2000.
XX
PF 02-DEC-1999; 99WO-EP09560.
XX
PR 07-DEC-1998; 98CB-0026886.
XX
PS (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI: 2000-423426/36.
DR N-PSDB; AAA48508.
XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
diagnostic, prophylactic and therapeutic purposes against microbial
diseases comprise a specific amino acid sequence
XX
PS Claim 4; Page 60-61; 98pp; English.
XX
CC The present sequence is a putative version of the Neisseria
meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
similar to the D15 outer membrane protein of the bacterium. The protein,
its gene, antibodies, antagonists and agonists can be used to diagnose
and treat bacterial diseases such as those leading to upper respiratory
tract infections, bacteraemia and meningitis. In addition, they can be
used in vaccines for use against cancer and autoimmune diseases.
XX
SQ Sequence 609 AA;
Query Match 100.0%; Score 609; DB 21; Length 609;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKPTALLPALFPFPFHAYAPAADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ 60
Db 1 MMLKPTALLPALFPFPFHAYAPAADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ 60
QY 61 DSEIKDMVEEHLPLITQOOEEVLDKEQTGF LAEBAPONVKTMRLSRGYSKVS/LTEKOG 120
Db 61 DSEIKDMVEEHLPLITQOOEEVLDKEQTGF LAEBAPONVKTMRLSRGYSKVS/LTEKOG 120
QY 121 AYTVHTTPGPRTKIANVGVAIIIGDILSDGNLAEEYRNALENWOOPVGSDFDQDSWENSKT 180
Db 121 AYTVHTTPGPRTKIANVGVAIIIGDILSDGNLAEEYRNALENWOOPVGSDFDQDSWENSKT 180
QY 181 SVLGAVTRKGYPLAKLGNTRAANVPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAANVPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
QY 241 SGLARFQPGTPTDLDLDFQALFQNGHYSGASVQADFRLQGDGRVPKVSYTEVKRHK 300
Db 241 SGLARFQPGTPTDLDLDFQALFQNGHYSGASVQADFRLQGDGRVPKVSYTEVKRHK 300
QY 301 LETGIRLDSEYGLGGKIAYDYNNLFNKGY IGSVYVMDMKYETTLLAAGISQPRNRYGNWYT 360
Db 301 LETGIRLDSEYGLGGKIAYDYNNLFNKGY IGSVYVMDMKYETTLLAAGISQPRNRYGNWYT 360
QY 361 SNVSYNRSTQNLKRAFGSGGIWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTQNLKRAFGSGGIWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
QY 421 LTASWKRRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
Db 421 LTASWKRRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
QY 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGSLE 540
Db 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGSLE 540
QY 541 YQLPFTRTLGSAGVAFHDMGDAANFKRMKLRHSGGLGYRVFSPFLAPFSFDIAYGHSDDKIR 600
Db 541 YQLPFTRTLGSAGVAFHDMGDAANFKRMKLRHSGGLGYRVFSPFLAPFSFDIAYGHSDDKIR 600
QY 601 WHISLGTFR 609
Db 601 WHISLGTFR 609
RESULT 3
AAY74862
ID AAY74862 standard; Protein; 615 AA.
AC AAY74862;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1198.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0034869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.

```
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) INST GENOMIC RES.
PA (GENO-)
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AA253624.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 669; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA254725 to AA2575941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254773 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 615 AA:
SQ
Query Match 40.9%; Score 249; DB 21; Length 615;
Best Local Similarity 99.6%; Pred. No. 4.7e-244;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MMKPTALLPALFFPHAYAPADLSKNAAGFALFNKNSPDTSVKLPKFPVRIDTQ 60
Db 7 MMKPTALLPALFFPHAYAPADLSKNAAGFALFNKNSPDTSVKLPKFPVRIDTQ 66
QY 61 DSEIKDWEELHLPLITQOQEVLDKQGTGFLAEEAPDNVKTMLRSKGYFSSKVSUTEKDG 120
Db 67 DSEIKDWEELHLPLITQOQEVLDKQGTGFLAEEAPDNVKTMLRSKGYFSSKVSUTEKDG 126
QY 121 AVTVHITPGPRTKIANYGVAITLGDILSDGNLAEYRYRNALENMQQPVGSDFDQDSWENSKT 180
Db 127 AVTVHITPGPRTKIANYGVAITLGDILSDGNLAEYRYRNALENMQQPVGSDFDQDSWENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNPOTATVDLNVVDSGRPTAFGDFEITGQRYPEQIV 240
Db 187 SVLGAVTRKGYPLAKLGNTRAAVNPOTATVDLNVVDSGRPTAFGDFEITGQRYPEQIV 246
QY 241 SGLARFQPGTPTDLDLLDFQQALEQNGHYSGASVQADFRLQGDVRPVKVSVTEVKRHK 300
Db 247 SGLARFQPGTPTDLDLLDFQQALEQNGHYSGASVQADFRLQGDVRPVKVSVTEVKRHK 306
QY 301 LETGRLDSEYGLGKKTAYDYNLFNKGYIGSVVWDMDKYETTLAAGISQPRNRYGNWT 360
Db 307 LETGRLDSEYGLGKKTAYDYNLFNKGYIGSVVWDMDKYETTLAAGISQPRNRYGNWT 366
QY 361 SNVSYNRSTTONLEKRAFSGGIWTVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSYNRSTTONLEKRAFSGGIWTVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 426
QY 421 LTASWKROLLNNVLPENGYLDGKIGTTLG 451
Db 427 LTASWKROLLNNVLPENGYLDGKIGTTLG 457
RESULT 4
AA259625
```

```
AA259625 standard; Protein; 587 AA.
AA259625;
03-JAN-2001 (first entry)
Neisseria meningitidis BASB040 putative protein sequence.
BASB040; bacterial disease; respiratory tract infection; bacteraemia;
meningitis; cancer; autoimmune disease.
Neisseria meningitidis.
WO200034480-A1.
15-JUN-2000.
02-DEC-1999; 99WO-EF09560.
07-DEC-1998; 98GB-0026886.
(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
Ruelle J;
WPI; 2000-423426/36.
N-PSDB; AAA48509.
Novel BASB040 polypeptides of Neisseria meningitidis useful for
diagnostic, prophylactic and therapeutic purposes against microbial
diseases comprise a specific amino acid sequence
Claim 4; Page 62; 98pp; English.
The present sequence is a putative version of the Neisseria
meningitidis strain H44/76 BASB040 protein sequence. This protein is
similar to the D15 outer membrane protein of the bacterium. The protein,
its gene, antibodies, antagonists and agonists can be used to diagnose
and treat bacterial diseases such as those leading to upper respiratory
tract infections, bacteraemia and meningitis. In addition, they can be
used in vaccines for use against cancer and autoimmune diseases.
SQ Sequence 587 AA:
Query Match 32.5%; Score 198; DB 21; Length 587;
Best Local Similarity 100.0%; Pred. No. 3.3e-192;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 412 DLGNSHATMLTASMKROLLNNVLPENGYLDGKIGTTLGTLTSLTALRTSARAGYFT 471
Db 390 DLGNSHATMLTASMKROLLNNVLPENGYLDGKIGTTLGTLTSLTALRTSARAGYFT 449
QY 472 PENKKLGTFIIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 531
Db 450 PENKKLGTFIIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 509
QY 532 RALLVGSLEYOLPRTTSLGAVFHDMDGDAANFRKMLKHGSLGVWFSPLAPSFDA 591
Db 510 RALLVGSLEYOLPRTTSLGAVFHDMDGDAANFRKMLKHGSLGVWFSPLAPSFDA 569
QY 592 YGHSDKKIRWHISLGRF 609
Db 570 YGHSDKKIRWHISLGRF 587
RESULT 5
AA259625
ID AA259625 standard; Protein; 615 AA.
XX
XX AA259625;
XX 21-MAR-2000 (first entry)
XX
```


|||||
Db 106 KTLRSKGYFSSKVSLEKDGAYTVHTPGPRTKIANVGVAIILGDILSDGNLAEEYRNAL 165
QY 160 ENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTNRAAVNPDAT 209
|||||
Db 166 ENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTNRAAVNPDAT 215
|||||

Search completed: November 9, 2002, 01:11:27
Job time : 36 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:10:43 ; Search time 20 Seconds
(without alignments)
895.928 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGSHDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: November 9, 2002, 01:14:22
Job time : 20 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:13:03 ; Search time 14 Seconds
(without alignments)
627.205 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMIKPTALLPALFFPPHAY.....IAYGHSDDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 92612 seqs, 14418503 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
---------------	----------------	--------------	----	-------------

No matches found

Search completed: November 9, 2002, 01:17:39
Job time : 14 secs

Qy 592 YGHSKKIRWHISLQTRF 609
 |||||
Db 618 YGHSKKIRWHISLQTRF 635

Search completed: November 9, 2002, 01:13:41
Job time : 21 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:03:58 : Search time 17 Seconds
(without alignments)
1485.828 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MMIKPTALLLPALFFPPHAY.....IAYGHSDKKIRWHISLGTRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	-------	-------------

No matches found

Search completed: November 9, 2002, 01:12:05
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 9, 2002, 01:09:24 ; Search time 33 seconds
(without alignments)
3802.506 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 609
Sequence: 1 MWIKPTALLPALFFPHAY.....IAYGSHDKIRWHISLGRF 609

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 671580 seqs, 206047115 residues

Word size : 15
Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_todent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_ivirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			ID	Description
	Score	Match	Length		
1	249	40.9	615	16 Q9JWM5	Q9Jwm5 neisseria m
2	198	32.5	635	16 Q9JXB7	Q9Jxb7 neisseria m

SUMMARIES

ALIGNMENTS

RESULT 1
Q9JWM5
ID Q9JWM5 PRELIMINARY; PRT; 615 AA.
AC Q9JWM5;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Putative outer membrane protein.
GN NMA0296.

RESULT 2
Q9JXB7
ID Q9JXB7 PRELIMINARY; PRT; 635 AA.
AC Q9JXB7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.

OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;

OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=63699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feitwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
RT *Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491*;
RL Nature 404:502-506(2000).
DR EMBL: AL162752; CAB83602.1; -
KW Complete proteome.
SQ SEQUENCE 615 AA; 67775 MW; B9A64607638E1CCB CRC64;

Query Match 40.9%; Score 249; DB 16; Length 615;
Best Local Similarity 99.6%; Pred. No. 2.6e-249;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MWIKPTALLPALFFPHAYAPAADLSENKAAGFALEKFNKSPDTSVKLKPKFPVRIDTQ 60
Db 7 MWIKPTALLPALFFPHAYAPAADLSENKAAGFALEKFNKSPDTSVKLKPKFPVRIDTQ 66
Qy 61 DSEIKDMVEEHLPLITQOOEEVLQKEQTGFLEAEAPDNVKTMLRSKGYFSSVSLTEKDG 120
Db 67 DSEIKDMVEEHLPLITQOOEEVLQKEQTGFLEAEAPDNVKTMLRSKGYFSSVSLTEKDG 126
Qy 121 AYTIVHTPGPRTKIANVGVAIIGDILSDGNAEYRYRLENWOOVCSDFDQDSWENSKT 180
Db 127 AYTIVHTPGPRTKIANVGVAIIGDILSDGNAEYRYRLENWOOVCSDFDQDSWENSKT 186
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVSDSGRPFAFGDFEITGTQRYPEQIV 240
Db 187 SVLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVSDSGRPFAFGDFEITGTQRYPEQIV 246
Qy 241 SGLAREQPGTPYDLDLLDQALQNGHYSGASVQADFQLOGDRVPVKVSVTEVKRHK 300
Db 247 SGLAREQPGTPYDLDLLDQALQNGHYSGASVQADFQLOGDRVPVKVSVTEVKRHK 306
Qy 301 LETGIRLDSYGLGGKIAYDYNNLFNKGYIGSVVYMDKYEYTTTAAAGISQPRNYRGNYWT 360
Db 307 LETGIRLDSYGLGGKIAYDYNNLFNKGYIGSVVYMDKYEYTTTAAAGISQPRNYRGNYWT 366
Qy 361 SNVSYNRSTQNLKRAFSGGTWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSYNRSTQNLKRAFSGGTWYVDRAGIDARLGAFFLAEGRKIPGSDIDLGNSHATM 426
Qy 421 LTASWKRQLLNVLHPENGHYLDGKIGTTLG 451
Db 427 LTASWKRQLLNVLHPENGHYLDGKIGTTLG 457

RESULT 2
Q9JXB7
ID Q9JXB7 PRELIMINARY; PRT; 635 AA.
AC Q9JXB7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein NMB2134.
GN NMB2134.

```

RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002561; AAF2442.1; -.
DR TIGR; NMB2134; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA; 69801 MW; 2B371F4A7A9DDA98 CRC64;

      Query Match      32.5%; Score 198; DB 16; Length 635;
      Best Local Similarity 100.0%; Pred. No. 2.5e-196;
      Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 DLGNSHATMLTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFT 471
Db 438 DLGNSHATMLTASWKRQLNNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFT 497
Qy 472 PENKLGTFIIRGQAGTYVARONADVPSGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 531
Db 498 PENKLGTFIIRGQAGTYVARONADVPSGLMFRSGGASSVRGYELDSIGLAGPNSVLPE 557
Qy 532 RALLVGSLEYQLPFTTSLGAVFHDMDGAAANFRMKLKHGSGLGVRWFSPPLAPFSFDIA 591
Db 558 RALLVGSLEYQLPFTTSLGAVFHDMDGAAANFRMKLKHGSGLGVRWFSPPLAPFSFDIA 617
Qy 592 YGHSDDKKIRWHISLGTRE 609
Db 618 YGHSDDKKIRWHISLGTRE 635
```

Search completed: November 9, 2002, 01:12:59
Job time : 33 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run On: November 9, 2002, 01:21:04 : Search time 266 Seconds
(without alignments)
5155.891 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPPHAY.....IAYGHSDKIRWHISLGRF 609

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delep 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: DEV=xlh
-MODEL=frame+_p2n.model
-Q=/cgn2.1/USPRO_pool/US09857669/runat_05112002_110007_7714/app_query.fasta_1.775
-DB=N_Geneseq_101002 -QPMF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPLC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669.scgn.1.145 @runat_05112002_110007_7714 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	3173	100.0	1830	21	AAA48507	Neisseria meningit
2	3173	100.0	1830	21	AAA48508	Neisseria meningit
3	3153	99.4	1848	21	AAZ53624	Neisseria meningit
4	3136	98.8	1848	21	AAZ53623	Neisseria meningit
5	3136	98.8	92934	21	AAA81473	N. meningitidis pa
6	3136	98.8	172325	21	AAF21613	Neisseria meningit
7	3019	95.1	1764	21	AAA48509	Neisseria meningit
8	2653	83.6	1569	21	AAZ53622	Neisseria gonorrhoe
9	384.5	12.1	1374	22	AAF61044	P. putida KT2440-a
10	374	11.8	1737	22	AAAF50269	Haemophilus influe
11	355.5	11.2	1731	21	AAA50270	Haemophilus influe
12	260.5	8.2	2503	21	AAA15155	DNA encoding outer
13	256.5	8.1	2379	22	AAZ507278	Neisseria meningit
14	256.5	8.1	2379	22	AAH42129	Nucleotide sequenc
15	253.5	8.0	2394	22	AAZ507279	Neisseria gonorrhoe
16	253.5	8.0	2394	22	AAH42130	Nucleotide sequenc
17	253	8.0	2394	22	AAZ507277	Neisseria meningit
18	253	8.0	2394	22	AAH42128	Nucleotide sequenc
19	253	8.0	52253	21	AAH81478	N. meningitidis pa
20	253	8.0	349980	21	AAAF21344	Neisseria meningit
21	251.5	7.9	2394	21	AAA15156	DNA encoding outer
22	247	7.8	6617	22	AAI97964	Lawsonia intracell
23	241	7.6	2193	23	AAZ53089	DNA encoding novel
24	234.5	7.4	2388	24	ABQ72980	Methylococcus caps
25	234.5	7.4	2391	24	ABO90024	M. capsulatus gene
26	231.5	7.3	1425	23	AAZ581897	DNA encoding novel
27	231.5	7.3	1425	23	AAZ53087	DNA encoding novel
28	219	6.9	2670	21	AAA50537	Moraxella catarrha
29	219	6.9	2760	21	AAA50536	Moraxella catarrha
30	219	6.9	89047	22	AAZ58547	Genomic fragment #
31	213.5	6.7	2989	15	AAQ66202	H. influenzae PAK
32	212.5	6.6	944	24	ABK52155	H. influenzae influe
33	210.5	6.6	2974	15	AAQ66201	H. influenzae SB33
34	209.5	6.6	2984	15	AAQ66199	H. influenzae b Ea
35	208.5	6.6	2953	15	AAQ66200	H. influenzae b Wl
36	207.5	6.5	2949	15	AAQ66198	H. influenzae b Ca
37	201.5	6.4	2019	22	AAAF61065	P. putida KT2440-a
38	174	5.5	2778	18	AAZ51338	H. pylori cell env
39	174	5.5	2778	20	AAZ51338	H. pylori outer me
40	173.5	5.5	2373	24	ABL91257	Chlamydia pneumonia
41	173.5	5.5	273254	21	AAZ51914	Chlamydia pneumonia
42	173.5	5.5	1230025	20	AAZ51990	Nucleotide sequenc
43	173	5.5	3012	19	AAZ52012	Helicobacter polyp
44	172.5	5.4	2751	22	AAZ52388	H. pylori HPS120 e
45	169.5	5.3	2550	21	AAZ51502	DNA encoding the C

ALIGNMENTS

RESULT 1
ID AAA48507 standard; DNA; 1830 BP.
XX
AC AAA48507;
XX
DT 03-JAN-2001 (first entry)
XX
DE Neisseria meningitidis BASB040 putative coding sequence #1.
XX
KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
XX
KW meningitis; cancer; autoimmune disease; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..1830
FT /tag. a

```

FT XX /product= "BASB040"
PN XX WO200034480-A1.
XX XX 15-JUN-2000.
XX XX 02-DEC-1999; 99WO-EP09560.
XX XX 07-DEC-1998; 98CB-0026886.
XX XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX XX Ruelle J;
PI XX WPI; 2000-423426/36.
DR DR P-PSDB; AAY99623.
XX XX
PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence
XX XX Claim 14; Page 59; 98pp; English.
XX XX
PS The present sequence is a putative version of the Neisseria
CC meningitidis strain ATCC13090 BASB040 coding sequence. The protein
CC produced from this gene is similar to the D15 outer membrane protein of
CC the bacterium. The gene, its protein, antibodies, antagonists and
CC agonists can be used to diagnose and treat bacterial diseases such as
CC those leading to upper respiratory tract infections, bacteraemia and
CC meningitis. In addition, they can be used in vaccines for use against
CC cancer and autoimmune diseases.
XX XX
SQ Sequence 1830 BP; 440 A; 603 C; 461 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 1,93e-285 Length: 1830
Score: 3173.00 Matches: 609
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-857-669-2 (1-609) x AAA48507 (1-1830)

QY 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20
DB 1 ATGATGATCAAAACCGACCGCCCTCTCTGCGGGCTTATTTTCTTCGCGACGCATAC 60
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
DB 61 GCGCCTCGCGCGACCTTTCGAAACAAGCGCGGGTTTCGATTGTTCAAAACAAA 120
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
DB 121 AGCCCGGACACCGAATCAGTTAAATTAACCCCAATTCCTCCGCGCATCGACACGCAG 180
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
DB 181 GATAGTGAATCAAAAGATATGTCGAAGAACACCTCGCGCTCATCAGCAGCAGGAA 240
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
DB 241 GAAGTATTGGACAAGGAACACACGGGCTCTCGCGGAAGAACACCGGCAACGTTAA 300
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
DB 301 ACCATGCTCCGACGAAAGGCTATTTCAGCAGCAAGATCAGCTGACGGAAAGACGGA 360
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
DB 361 GCTTATACGGGTACACATCACACCGGGCGCGCACCAAAATCGCCAACTCGGTGTCGCC 420
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160

```

```

DB 421 ATCTCTCGCGGACATCTCTTTCAGACGGCAACCTCGCCGAATACTACCGCAACGGCTGGAA 480
QY 161 AsnTrpGlnGlnProValIglySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
DB 481 AACTGSCAGCAGCGGTAGGAGTATTCGATCAGGACAGTTGGGAACAACAGCAAACT 540
QY 181 SerValIleuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
DB 541 TCGGTCTCTCGCGCGGTAAACGGAAAGGCTACCGGTTCCCAAGCTCGCCACACCCGG 600
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
DB 601 GCGGCGGTCAACCGCGATACGCCACCGCTCGATTGTAACGTCGTCGTGGACAGCGCGC 660
QY 221 ProIleAlaPheGlyAspPheIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
DB 661 CCCATCGCCTTCGCGCACTTGAATATCACCGCACAGCGTTACCCCGAACAATCGTC 720
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
DB 721 TCGGCGCTGCGCGCTTCCAAACCGGCGACGCCCTACGACCTCGACCTGCTCGACTTC 780
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
DB 781 CAACAGGCACTCAACAACAAACGGGCAATTATTCGCGCGCTCGCTACAGCCGACTTCGAC 840
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
DB 841 CGTCTCAAGGCGACGGGCTCGCGTCAAGTCAGCTAACCGAGGTAAACGCGCAACGCAAG 900
QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
DB 901 CTCGAACCGCGCATCTCGCTCGATTGGAATACGGTTTGGCGGCAAAATCGCTACGAC 960
QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
DB 961 TATTACACCTCTTCAACAAGGCTATATCGGCTCGGTCGCTCGGATATGGCAAAATAC 1020
QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
DB 1021 GAACCCAGCTTTCGCGCGGCGCATCAGCGCGCGCAACTATCGGGCGCACTACTGGACA 1080
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
DB 1081 AGCAAGTTTCTTACACACGTTTCGACACCGCTCGAACCCCAAACTCGAAACCGCGCTTC 1140
QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
DB 1141 GGCATCTGCTATGTCGCGACCGCGCGGCGCATCGATCGCAGGCTGGGGGCGAGTTTCTC 1200
QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
DB 1201 GCAGAAGCGCGGAAATCCCGGCTCGGATATCGATTGGGCAACAGCCACGCCAGATG 1260
QY 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
DB 1261 CTGACCGCTCTTGGAAACCGCAGCTCTCAACAACGTCGTCGATCCCGAAGAGGCCAT 1320
QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
DB 1321 TACTCTGCGCGGAAATCGGACGACTTTGGGCACATTCCTGCTCCACCGCGCTGATC 1380
QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
DB 1381 CGCACCTCTGCGCGTGCAAGTTATTCTTCACGCCCGCAAAACAAACAACTCGGCACG 1440
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
DB 1441 ATCATACCGCGGACAGCGGGTTACACCGTTCGCGCGCAATCGCGACGTTCTTCAGGG 1500
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520

```


Db 1501 CTGATGTTCCGCGAGCGCGCGTCTCCGTCCGCGGTACGAACCTCAGACGATCGGA 1560

Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540

Db 1561 CTTGCCGCGCCGAAACGATCGTCTGCCGAAACGCGCTCTCTGTGGCGAGCCTGGAA 1620

Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560

Db 1621 TACCAACTGCCGTTTACCGGACACCTTTCGCGCGCGGTGTCCACGATATGGCGATGCC 1680

Qy 561 AlaAlaAsnPheGlyArgMetLeuLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580

Db 1681 GCGCCCAATTTCAACAGTATGAAGCTGAACACAGGTTCGGGACTGGCGCTGGGTTC 1740

Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600

Db 1741 AGCCCGCTTTCGCGCGCTTTCCTTCGACATCGCTACGGGCACAGCGATAGAAATCCGC 1800

Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609

Db 1801 TGGCAGATCAGCTTGGGAACACGCTTC 1827

RESULT 2

AAA48508

ID AAA48508 standard; DNA; 1830 BP.

XX

AC AAA48508;

XX

DT 03-JAN-2001 (first entry)

XX

DE Neisseria meningitidis BASB040 putative coding sequence #2.

XX

KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;

KW meningitis; cancer; autoimmune disease; ds.

XX

OS Neisseria meningitidis.

XX

FH Key Location/Qualifiers

FT CDS 1..1830

FT /*tag= a

FT /product= "BASB040"

XX

PN WO200034480-A1.

XX

XX 15-JUN-2000.

XX

XX 02-DEC-1999; 99WO-EP09560.

XX

XX 07-DEC-1998; 98GB-0026886.

XX

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

XX Ruelle J;

XX

XX WPI; 2000-423426/36.

XX

XX P-PSDB; AAY99624.

XX

XX Novel BASB040 polypeptides of Neisseria meningitidis useful for

XX diagnostic, prophylactic and therapeutic purposes against microbial

XX diseases comprise a specific amino acid sequence

XX

XX Claim 11; Page 60; 98pp; English.

XX

XX The present sequence is a putative version of the Neisseria

XX meningitidis strain ATCC13090 BASB040 coding sequence. The protein

XX produced from this gene is similar to the D15 outer membrane protein of

XX the bacterium. The gene, its protein, antibodies, antagonists and

XX agonists can be used to diagnose and treat bacterial diseases such as

XX those leading to upper respiratory tract infections, bacteraemia and

XX meningitis. In addition, they can be used in vaccines for use against

XX cancer and autoimmune diseases.

XX

XX Sequence 1830 BP; 439 A; 604 C; 461 G; 326 T; 0 other;

XX

Alignment Scores:

Prod. No.: 1.93e-285 Length: 1830

Score: 3173.00 Matches: 609

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 21 Gaps: 0

US-09-857-669-2 (1-609) x AAA48508 (1-1830)

Qy 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20

Db 1 ATGATGATCAAAACCCAGCGCGCTGCTCTCGCGCTTTATTTTCTTCGCGACGCATAC 60

Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40

Db 61 GGCCTCGCCCGCGACTTTCGAAACAAAGCGCGGGTTCGCATTGTTCAAAACAAA 120

Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60

Db 121 AGCCCGCACCGGAATCAGTTAAATTAACCCCAATTCGCCGTCCGCATCGACACGCAG 180

Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80

Db 181 GATAGTGAATCAAGATATGTCGAGAACACCTGCCGTCTATCAGCGACGACGAGAA 240

Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100

Db 241 GAAGTATTGGACAGGAACAGACGGGCTTCCTCGCCGAAGAAGCACGGACAACGTTAAA 300

Qy 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120

Db 301 ACCATGCTCCGCGCAAGAGGTATTTTCAGCAGCAAGTACGCTGACGGAAGACGGA 360

Qy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140

Db 361 GCTTATACGCTACATCACACCGCGCGCGCACCAAAAATCGCCAACGTCGCTGCGCC 420

Qy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160

Db 421 ATCTCGCGGCACATCTTTTCAGACGGCAACCTCCGCCAANTACTACCGCAACGCGCTGAAA 480

Qy 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180

Db 481 AACTGCGCAGCAGCGCTAGGCAGTGATTTTCGATCAGGACAGTTGGGAAAACAGCAAACT 540

Qy 181 SerValIleGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200

Db 541 TCGTCTCTCGCGCGCGGTAAACGCGCAAGGCTACCGCTTCCCAAGCTCGGCAACACCCGG 600

Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220

Db 601 GCGCGCGTCAACCCGATACCGCCACCGCTGATTTGAACGTCGCTGCGGACAGCGCGCG 660

Qy 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240

Db 661 CCATCGCTCTCGCGGACTTTGAATACACCGGCACACAGCTTACCCCGCAACAAATCGTC 720

Qy 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260

Db 721 TCGCGCTCGCGCGCTTCAACCGGCGACCGCTACGACCTCGACCTGCTGCTCGACTTC 780

Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280

Db 781 CAACAGCAGCTCGAACAACCGGCGATTAATTCGCGCGCTCGCTTACAAAGCGGACTTCGAC 840

Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300

Db 841 CGTCTCAAGGCGCGCGCTCCCGCTCAAGTCAAGCTAAGGCTAAGGCTCAAGGCTCAAGGCTCAAG 900

Qy 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320

Db 901 CTCGAACCGGCATCCGCTCGATTGCGAATACGGTTTGGCGCGCAAAATCGCCTACGAC 960
 Qy 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
 Db 961 TATTACAACTCTTCAACAAAGCTATATCGCTCGGTCTGGGATATGGACAAATAC 1020
 Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 Db 1021 GAAACCAAGCTTCCGCGCGCATACGCGAGCGGCAACTATCGGGCGCAACTACTGSGACA 1080
 Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
 Db 1081 AGCAAGCTTTCCTACAAACGCTTCGACACCCCAACAAACCTCGAAAAACGGCTTCTCCGGC 1140
 Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
 Db 1141 GGCATCTGGTATCTGCGCACCGCGCGGCATCGATCGACGCTGGGGCAGAGTTTCTC 1200
 Qy 401 AlaGluIArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
 Db 1201 GCAGAGCGCGGAAATCCCGGCTCGGATATCGATTGGGCAACACCGCCACGCGCATG 1260
 Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis 440
 Db 1261 CTCACGCGCTCTTGGAAACGCGCAGCTGTCTCAACAAACGCTGTGCATCCCGAAACGGCCAT 1320
 Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
 Db 1321 TACTCGACGGCAAAATCGTACGACTTTGGGCACATTCCTCTCCACGGCGCTGATC 1380
 Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
 Db 1381 CGCACCTCTCCCGCTCGAGGTATTCTTCACGCGCGGAAACAAACAACTCGGCACGTTTC 1440
 Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
 Db 1441 ATCATACGGGCAACGGGTATACACCGTTGCCCGCAATGCCGACGTTCCTTCAGGG 1500
 Qy 501 LeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
 Db 1501 CTGATGTTCGCGACGCGCGCGCTCTCCGTGCGCGGTTACGAACTCGACGATCGGA 1560
 Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
 Db 1561 CTTCGCGCGCGAAGCATCGGTCTGCCGACGCGCCCTCTGTGTGGCAGCCTGGAA 1620
 Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
 Db 1621 TAGCAACTGCGGTTTACGGCGACCCCTTTCCGCGCGGTTTCCACGATATGGCGGATGCC 1680
 Qy 561 AlaAlaAsnPhelLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
 Db 1681 GCGCCCAATTCACAGCTATGAAGCTGAACACAGGTTGCGGACTGGCGGTGCGCTGGTTC 1740
 Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
 Db 1741 AGCCGCGTTGCGCGGTTTTTCTTCGACATCGCTACGGGCACAGCGATAAGAAATCCGC 1800
 Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
 Db 1801 TGGCACAATCAGCTTGGGACACGCTTTC 1827

RESULT 3

AAZ53624 standard; DNA; 1848 BP.

XX AAZ53624;

XX 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 286 partial DNA sequence SEQ ID NO:1197.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

antigenic; diagnosis: immunogenic; infection; meningitis; septicaemia;
antibacterial; gene therapy; ds.

Neisseria meningitidis.

WO9957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

02-SEP-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

09-OCT-1998; 98US-0103796.

23-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
Tettelin H, Venter JC;

WPI: 2000-062150/05.

P-PSDB: AAY74862.

Novel Neisserial polypeptides predicted to be useful antigens for
vaccines and diagnostics

Claim 7; Page 668-669; 1453pp; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
presence of Neisseria bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

SQ Sequence 1848 BP; 440 A; 614 C; 468 G; 326 T; 0 other;

Alignment Scores:

Pred. No.: 1,42e-283 Length: 1848
 Score: 3153.00 Matches: 605
 Percent Similarity: 99.51% Conservative: 1
 Best Local Similarity: 99.34% Mismatches: 3
 Query Match: 99.37% Indels: 0
 DB: 21 Gaps: 0

US-09-857-669-2 (1-609) x AAZ53624 (1-1848)

Qy 1 MetMetIleLysProThrAlaLeuLeuLeuProAlaLeuPhePheProHisAlaTyr 20

Db 19 ATGATGATTAAACCGACCGCCCTGCTCTCGCGGCTTTATTTCTTCGCGACGATAC 78

Qy 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40

Db 79 GCGCCTGCCCGCACCTTTCCGAAACAAAGCGGGGTTTCGCATTGTTCAAAACAAA 138

Qy 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60

Db 139 AGCCCCGACACCGAATCAGTTAAATTAACCAACCAATTTCCCGCTCCGATCGACGACG 198

61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
199 GATAGTGAATCAAGATATGTCGAAGAACAACCTGCCCTCATCACGACGACGAGAA 258
81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
259 GAAGTATTGGACAAGGAACAACACGCGGCTTCCCTCGCGAAGAACACCGGACAACTTAA 318
101 ThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeuThrCluLysAspGly 120
319 ACAATGCTCCGACGAAGGCTATTTCAGCAGCAAGTACGCTGACGGAAGAACACGGA 378
121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
379 GCTTATAGCGTACACATCACACCGCGCGCCGACCAAAATGCCAACGTCGCGCTGCC 438
141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
439 ATCCCTCGCGGACATCCTTTACAGCGCAACCTCGCGAATACCTACCGCAACCGCTGGA 498
161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
499 AACTGGCAGCAGCGGAGGAGGAGTATTCGATCAGGACAGTGGGAAAACAGCAAACT 558
181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
559 TCCGCTCTCGCGGCGGTAAACCGCAAGCCCTACCGCTTGCAGGCTCGGCAACACCGG 618
201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
619 CGCGCGCTCAACCCCGATACCGCCACCGCGGATTTGAACGTCGCTGGACAGCGCGC 678
221 ProIleAlaPheGlyAspPheGluIleThrGlyThrClnArgTyrProGluGlnIleVal 240
679 CCCATCGCTTCGGGGACTTTGAANTTACCGGACGAGCGGTTACCCCGAACAATCGTC 738
241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
739 TCCGCGCTTGGCGCGCTTCAACCGCGGACCGCCCTACGACCTCGACCTGCTCGACTTC 798
261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
799 CAACAGGCGCTCGAACAACAAACGCGCATTTATTCGGCGCGCTCCGTACACCGCATTCGAC 858
281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
859 CGCCTCCAAGCGCGCGCTCCCGTCAAAAGTCAGCGTAACCGAGGTCAACGCGCAAG 918
301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp 320
919 CTCGAACCGGATCCGCTCGATTCGGAATACGTTTGGCGGCAAAATCCCTACGAC 978
321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
979 TATTACAACTCTTCAACAAAGGCTATATCGGTTCCGTCGCTCGGATATGGCAAAATAC 1038
341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
1039 GAACACGCTTCCCGCGCATCGCCAGCGGCACTATCGGCGCAACTACTGGACA 1098
361 SerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPheSerGly 380
1099 AGCAACGTTTCTACACCGGTTTCGACACCAACCAACCTCGAAAAACCGCCTCTCCGCG 1158
381 GlyIleTrpTyrValArgAspAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
1159 GGCATCTGGTATGTCGGGACCGCGCGGGCATCGATCCAGGCTGGGGCGGAGTTCTC 1218
401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
1219 GCAGAAGCGCGGAAATCCCGCGCTCGGATATCGATTGGGCAACACGACCGCCACGATG 1278
421 LeuThrAlaSerTrpLysArgGlnLeuAsnAsnValLeuHisProGluAsnGlyHis 440

1279 CTGACCGGCTCTTGAACGCGCAGCTGCTCAACACGTTGCTGCATCCCGAAGCGCCAT 1338
441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
1339 TACCTCGACGCGCAAAATCGGTACGACTTTGGCGCATTTCTGTCCTCCACCGCGTGATC 1398
461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
1399 CCACCTCTGCGCGGCGAGGTATTTCTTCCGCCCGAACAACAAAATTCGCGACGTTTC 1458
481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
1459 ATCATACGCGACAAGCGGGTTACACCGTTGCCGCGACAAATGCCAACCTTCTTCAGGG 1518
501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
1519 CTGATGTTCCGCGCGCGCGCGCTTCTCGTGGCGGTTACGAATCCGACGATCGGG 1578
521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
1579 CTTGCGCGCGCGCAACGATCGGTCTGCCGAACGCGCCCTCTTGGTGGCGACGCTGGA 1638
541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
1639 TACCAACTCGCGCTTTACGCGCACCCCTTTCGCGCGGCTGTCCACGATATGGCGACGCC 1698
561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
1699 GCGCGCAATTTCAACGCTATCAAGCTGAACACGCTTCGGGACTGGCGCTGGCTGTTTC 1758
581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
1759 AGCCGCTCGCGCGCTTTCTTCCGATCGCTTACGCGCGGCTGTCCACGATATGGCGACGCC 1818
601 TrpHisIleSerLeuGlyThrArgPhe 609
1819 TGGCACATCAGCTTGGGAACGCGCTTC 1845
RESULT 4
ID AA253623 standard; DNA; 1848 BP.
XX AA253623;
DT - 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 286 partial DNA sequence SEQ ID NO:1195.
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX Neisseria meningitidis.
OS
XX WO957280-A2.
PD 11-NOV-1999.
XX
PF 30-APR-1999; 95WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 98US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX

QY 541 TvrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
 DB 1639 TACCACACTGCGGTTTACCGCGACCCCTTCCGGCGGGTGTCCACGATATGGCGCATGCC 1698
 QY 561 AlaAlaAsnPhelYsArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
 DB 1699 GCGGCAATTTCAACACGTATGAAGCTGAACACGGTTCCGGACTGGCGCTGGCTGTTTC 1758
 QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
 DB 1759 AGCGCGCTTGGCGGCTTTTCCTTCGACATCGCTACGGCACACGGATGAAGAAATCCGCG 1818
 QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
 DB 1819 TGGCACATCAGCTTGGGAACGCGCTTC 1845
 RESULT 5
 ID AAA81473 standard; DNA; 92934 BP.
 AC AAA81473;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_21 SEQ ID NO:21.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7; Page 471-498; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 92934 BP; 24051 A; 24249 C; 21902 G; 22730 T; 2 other;
 Alignment Scores:
 Pred. No.: 1.14e-279 Length: 92934
 Score: 3136.00 Matches: 601
 Percent Similarity: 99.18% Conservative: 3
 Best Local Similarity: 98.69% Mismatches: 5
 Query Match: 98.83% Indels: 0
 DB: 21 Gaps: 0
 US-09-857-669-2 (1-609) x AAA81473 (1-92934)
 QY 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
 DB 38268 ATGATGATCAAAACCGACCGCCTCTCTCGCGCTTTATTTTCTTCCGACGCATAC 38327
 QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
 DB 38328 GCGCCTGCGCGCGACCTTTCGNAACAAAGCGCGGGTTCGCATTGTTCAAAACAAA 38387
 QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
 DB 38388 AGCCCGCAGCACGGAATCAGTCAAAATTAACACCCCAATTCCTCTCATCGACACGCG 38447
 QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
 DB 38448 GACAGTGAATCAAGATATGTCGAAGAACACCTGCGCGCTCATACGACGACGAGGAA 38507
 QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
 DB 38508 GAAGTATTGGACAAGGAACAGACGCGCTTCTCGCGCAAGACGCGGACAACTTAA 38567
 QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
 DB 38568 AGATGCTCCGCGCAGCAAGGCTATTTTCAGCAGCAAGCTGACCTGACGGAAGAAGCGA 38627
 QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
 DB 38628 GCTTATACGGTACATCACACCGCGCGCGCCACCAAAAATCGCCAACTCGCGCTCGC 38687
 QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
 DB 38688 ATCCTCGCGCAGCATCTCTTCAGACGGCAACCTCGCCGAANTACTACCGAACCGCTG 38747
 QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
 DB 38748 AACTGCGCAGCAGCGGTAGGCGAGCGATTCGATCAGCACAGTTCGGAAAAACAGCAA 38807
 QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
 DB 38808 TCGTCTCTCGCGCGCGGTAAACCGGCAACCTTACCCCGTTCCGAAGCTCGCGCAATAC 38867
 QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
 DB 38868 GCGGCGCTCAACCCCGATACCGCCACCGCGATTTGAACGCTCGTGGACAGCGCGCCG 38927
 QY 221 ProIleAlaPheGlyAspPheCulThrGlyThrGlnArgTyrProGluGlnIleVal 240
 DB 38928 CCATCGCCTTCGCGGAGCTTTGAATACACCGGCACACAGGCTTACCCCGAACAAATCGTC 38987
 QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
 DB 38988 TCGGCGCTTTCGCGCTTTCAGCCCGGTATGCGCGTACGACCTCGACCTGCTGCTCGACTTC 39047
 QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280

```

Db 39048 CAACAGCGCTCGAACAACGGCGATTATTCGGCGCGTCCGTACAGCGCGACTTCGAC 39107
Qy 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 39108 CGGCTCCAGAGGCGCGCTCCCGTCAAGATCAGCTACCGAGGTCAACAGCCCAAA 39167
Qy 301 LeuGluThrGlyIleArgLeuAspSerGluThrGlyLeuGlyGlyLysIleAlaThrAsp 320
Db 39168 CTCGAACCGCATCGCTCGATTCCGAATACGGTTGGCGGCGCAAAATCGCTACGAC 39227
Qy 321 TyrThrAsnLeuPheAsnLysGlyTyrIleGlySerValValThrAspMetAspLysThr 340
Db 39228 TATTACAACCTCTTCAACAAGAGCTATATCGGTTCCGTCGTGGGATATGGACAAATAC 39287
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnThrArgGlyAsnThrTrpThr 360
Db 39288 GAAACACAGCTTGGCGCGGCATCAGCCAGCGCGCAACTATCGGGCGCAACTACTGACA 39347
Qy 361 SerAsnValSerThrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 39348 AGCAACGTTTCTTACAAACGGTTCGACCAACCCCAAACTCGGAAACCGCGCTTCTCCGGC 39407
Qy 381 GlyIleThrThrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 39408 GCGCTCTGCTATGTGCGCGACCGCGCGGCATCGATCCAGGCTGGGGCGGAAATTTCTC 39467
Qy 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 39468 GCAGAAGCGGGAATATCCCGGCTCGGCTCGATTTGGGCAACAGCCACGCGCATG 39527
Qy 421 LeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
Db 39528 CTGACCGCTCTTGGAAACGCGAGCTGCTCAACAACGTGTCATCCGAAACAGGCCAT 39587
Qy 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerThrAlaLeuIle 460
Db 39588 TACCTCGACGGCAAAATGGTACGACTTTGGGCACATTCCTGCTCCACCGCGCTGATC 39647
Qy 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 39648 CGCACCTCTGCGCGTGCAGGTATTCTTCACGCCCGCAAAACAACTCGGCACGTTTC 39707
Qy 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 39708 ATCATACGGCGGACAGCGGTTCACCCCTGCCCGCGCACATGCCGACGTTCTTCAGGG 39767
Qy 501 LeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 39768 CTGATGTTCCGACGGCGGCGCGCTCTTCGTCGCGGTTCAGAACTCGACAGCATCGGA 39827
Qy 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 39828 CTTGCGCGGCGCAACGGATCGGTCTGCGCGAAGCGCCCTCTCTGTGGCGACGCTGGAA 39887
Qy 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 39888 TACCAACTCGCGCTTACGCGCACCGCTTTCGCGCGCGGTGTTCACCATATATGGCGATGC 39947
Qy 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 39948 GCGCGCAATTTCAACGATATGAAGCTGAACACGGTTTCGGGACTGCGGCTGCGGTTC 40007
Qy 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 40008 AGCGCGCTTGGCGCGCTTTTCTCTCGACATCGCTACGGGACAGCGATAGAATAATCCGC 40067
Qy 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 40068 TGGCACATCAGCTTGGGAACGCGCTTC 40094

```

RESULT 6
AAF21613

AAF21613 standard; DNA; 172325 BP.
AAF21613;
13-MAR-2001 (first entry)
Neisseria meningitidis B nucleotide sequence SEQ ID NO:114.
Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.
Neisseria meningitidis.
WO2000066791-A1.
09-NOV-2000.
08-MAR-2000; 2000WO-US05928.
30-APR-1999; 99US-0132068.
08-OCT-1999; 99WO-US23573.
28-FEB-2000; 2000GB-0004695.
(CHIR) CHIRON CORP.
(GENO-) INST GENOMIC RES.
Pizza M, Hickey B, Peterson J, Tettelin H, Venter JC, Maignani V,
Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
Frazer CM, Grandi G;
WPI; 2000-647603/62.
Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
Claim 7: Appendix A: 692pp; English.
The present invention describes the full length genome of
Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
to AAF21613 represent fragments of the NMB genomic sequence, as the
sequence was too long to go in a record on its own it was split into 8
sequences which overlap each other at the beginning and end of each
sequence by 4980 bp (i.e. the last 4980 bp of AAF21544 is repeated at
the beginning of AAF21607, the last 4980 bp of AAF21607 are repeated at
the beginning of AAF21608, and so on). AAF21545 to AAF21589 encode the
Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
AAF21606 represent PCR primers which are used in the exemplification of
the present invention. The NMB genome and fragments from it have
antibacterial activity, and can be used in vaccines and gene therapy.
Neisseria nucleic acids, proteins and/or antibodies which binds to the
proteins can be used in compositions for treating or preventing infection
due to Neisserial bacteria or as a diagnostic reagent for detecting the
presence of Neisserial bacteria or of antibodies raised to Neisserial
bacteria. Computers, computer memory, computer storage medium or computer
databases can be used in a search to identify open reading frames (ORFs)
or coding sequences within the NMB genome. The DNA sequences provide
further opportunities to find antigenic or immunogenic proteins which are
more effective in vaccines than the outer membrane proteins currently
used.
SQ Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;
Alignment Scores:
Pred. No.: 2.66e-279 Length: 172325
Score: 3136.00 Matches: 601
Percent Similarity: 99.18% Conservative: 3
Best Local Similarity: 98.69% Mismatches: 5
Query Match: 98.83% Indels: 0
DB: 21 Gaps: 0
US-09-857-669-2 (1-609); x AAF21613 (1-172325)

```
QY 1 MetMetIleLysProThrAlaLeuLeuProAlaIleuPhePheProHisAlaTyr 20
Db 141935 ATGATGATCAAMCCGACCGCCTCTCTCCGGCGTTATTTTCTTCCGACGCATAC 141994
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 141995 GCGCCTGCGCGCGACCTTTCGAAAAACAAGCGCGCGGTTTCGATTGTTCAAAAACAA 142054
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 142055 AGCCCGACACCGATCATAGTCMAATTAAACCCCAATTCGCCGTCCTCATCGACACGGAG 142114
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 142115 GACAGTGAATAACAAGATATGGTCCAAAGAACACTGCCGCTCATCGCAGCAGCAGGAA 142174
QY 81 GluValLeuAspLysGluGlnThrGlyPheIleuAlaGluAlaProAspAsnValLys 100
Db 142175 GAAGTATTGGACAGGAACACGCGGCTTCTGCCGAGAGAGCGCGGCAACCGTTAA 142234
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeuThrGluLysAspGly 120
Db 142235 AGCATGCTCCGCGACAAAGGCTATTTCAGCAGCAAAAGTCAGCCTGACGGAAGACGGA 142294
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyAla 140
Db 142295 GCTTATACGGTACATCACACCGCGCGCCGACCAAAATCCCAACGTCGCGCTCGCC 142354
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 142355 ATCCTCGCGGACATCTTTCAGACGGCAACTCGCCGAATACTACGCCAACGCGTGGAA 142414
QY 161 AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 142415 AACTGACACGACCGCGGTAGCAGCGATTTTCATCAGGACAGTTGGGAAACAGCAAACT 142474
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 142475 TCCGTCCTCGCGCGCGGTAAACGCGAAAGCCTTACCGCTTGCCCAAGCTCGGCAATACG 142534
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 142535 GCGGCGGTCAACCCCGATACCGCCACCGCGGATTTGAACGTCGCTGGACAGCGCGCG 142594
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 142595 CCCATCGCCTTCGGCGACTTTGAATACCGGCACACAGCGTTACCCCGAACAATTCGTC 142654
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 142655 TCCGCGCTTCCCGTTTCCACCGCGGTATGCCGTACGACCTCGACCTGCTCGACTTC 142714
QY 261 GlnGlnAlaLeuGlnAlaAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 142715 CAACAGGCGCTCGAAACAACAGCGGATTAATTCGCGCGCTCCGCTACAAAGCGACTTCGAC 142774
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 142775 CGCCTCCAAGGCGACCGGTCCTCGGTCAAGTCAGCGTAAACCGAGGTCAAAACCCCAAA 142834
QY 301 LeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAsp 320
Db 142835 CTCGAACCGGATCCGCTCGATTTCGGAATACGTTTGGCGGCAAAATTCGCTACGAC 142894
QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 142895 TATTACACCTCTTCAACAAGGCTATATCGGTTCCGTCGTGGGATATGGCAANTAC 142954
QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 142955 GAAACACGCTTTCGCGCGGATCAGCCAGCGGCAACTATCGGGCAACTACTGGACA 143014
QY 361 SerAsnValSerAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
```

```
Db 143015 AGCAACGTTTCTTACACCGTTCCACCACCCAAACCTCGAAACACGCCCTTCTCCGCG 143074
QY 381 GlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeu 400
Db 143075 GCGCTCTGTATGTGCGCGACCGCGCGGCATCGATGCCAGCTGGGGCGGAATTTCTC 143134
QY 401 AlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMet 420
Db 143135 GCAGAAGCCGGAANAATCCCGCGCTCGCTGTGATTTGGGCAACAGCCACCGCATG 143194
QY 421 LeuThrAlaSerTyrPlysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHis 440
Db 143195 CTGACCGCTCTCTTGAAACGCGCAGCTGCTCAACAACGCTGTGCTATCCGAAAACGCCAT 143254
QY 441 TyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIle 460
Db 143255 TACCTCGACGGCAAAATCGTACGACCTTTGGGCACATTTCTGTCTCCACCGCGCTGATC 143314
QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPhe 480
Db 143315 CGCACCTCTGCCCGTGCAGGTTATTTCTTTCACGCCGCAAAACCAAAACCTCGGCACGTTTC 143374
QY 481 IleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGly 500
Db 143375 ATCATACGGGCAACAGCGGTTTACCGTTGCCGCGCACTGCCGACGTTCTCTCAGGG 143434
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 143435 CTGATGTTCCGACGCGCGCGGCTCTTCGTCGCGGTTTACGAACTCGACAGCATCGGA 143494
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 143495 CTTGCGGCGCCGACGATCGGTCTCGCCGACGCGCCCTCTCTGTTGGGACGCTGGAA 143554
QY 541 TyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 560
Db 143555 TACCAACTCGCGTTTACGCGCACCGCTTTCGCGCGCGGTGTTCACGATATGGCGATGCC 143614
QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 143615 GCGGCCAATTTCAACGATGTAAGCTGAAACACGCTTCGGGACGTTGGCGTGGCTGGTTC 143674
QY 581 SerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArg 600
Db 143675 AGCCCGCTTGGCGCGCTTTTCTTCGACATCGCTACGGCACAGCGCATAGAAAATCCGC 143734
QY 601 TrpHisIleSerLeuGlyThrArgPhe 609
Db 143735 TGGCACATCAGCTTGGGAACGCGGCTTC 143761
RESULT 7
ID AAA48509 standard; DNA; 1764 BP.
XX AAA48509;
XX AAA48509;
XX 03-JAN-2001 (first entry)
XX
XX Neisseria meningitidis BASB040 putative coding sequence.
XX
XX BASB040; bacterial disease; respiratory tract infection; bacteraemia;
XX meningitis; cancer; autoimmune disease; ds.
XX
XX Neisseria meningitidis.
XX
XX Key Location/Qualifiers
XX CDS 1..1764
XX FT /*tag= a
XX FT /product= "BASB040"
XX FT /partial
XX
XX W0200034480-A1.
```



```

XX 15-JUN-2000..
PD
XX
XX 02-DEC-1999: 99WO-EP09560.
PF
XX 07-DEC-1998: 98GB-0026886.
PR
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX
XX Ruelle J;
PI
XX
XX WPI: 2000-423426/36.
DR
XX P-PSDB: AAY99625.
XX
XX Novel BASB040 polypeptides of Neisseria meningitidis useful for
PT diagnostic, prophylactic and therapeutic purposes against microbial
PT diseases comprise a specific amino acid sequence -
XX
XX Claim 11: Page 61; 98pp; English.
PS
XX
XX The present sequence is a putative version of the Neisseria
CC meningitidis strain H44/76 BASB040 coding sequence. The protein
CC produced from this gene is similar to the D15 outer membrane protein of
CC the bacterium. The gene, its protein, antibodies, antagonists and
CC agonists can be used to diagnose and treat bacterial diseases such as
CC those leading to upper respiratory tract infections, bacteraemia and
CC meningitis. In addition, they can be used in vaccines for use against
CC cancer and autoimmune diseases.
XX
XX Sequence 1764 BP: 424 A: 581 C: 451 G: 308 T: 0 other:
SQ

```

```

Alignment Scores:
Pred. No.: 3,93e-271 Length: 1764
Score: 3019.00 Matches: 579
Percent Similarity: 99.15% Conservative: 3
Best Local Similarity: 98.64% Mismatches: 5
Query Match: 95.15% Indels: 0
DB: 21 Gaps: 0

```

```

US-09-857-669-2 (1-609) x AAA48509 (1-1764)
QY 23 AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerPro 42
DB 1 GCCGCCGACCTTCCGAAACAAAGCGCGGGTTCCGATGTTCAAAAACAAACCCGCC 60
QY 43 AspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSer 62
DB 61 GACACCGAATCAGTCAAAATTAAACCCCAAAATTCGCCGCTCATCGACACGACGACGT 120
QY 63 GluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGluVal 82
DB 121 GAAATCAAGATATGTCGAGAACACCTGCCGCTCATCAGCAGCAGCAGGAGGA 180
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
DB 181 TTGGACAAGGAACACACCGGCTTCCTCCGCGAAGACGCGGACAACTTAAACGATG 240
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
DB 241 CTCGCCAGCAAAAGGCTATTTCAGCAGCAAAAGTCAGCTGACGGAAGAACGCGCTTAT 300
QY 123 ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
DB 301 ACGGTACACATCACACCGGGCCGCGCACCAAAAATCCCAACCTCGCGCGTCCCTC 360
QY 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
DB 361 GCGGACATCCTTTCAGACGGCAACCTCCGCGAATACTACCGCAACGCGCTGGAAACTGG 420
QY 163 GlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerVal 182
DB 421 CAGCAGCCGGTAGGGAGCGGATTTTCATCAGGACAGTGTGGGAAACACGACAAACCTCCGTC 480

```

```

183 IeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAla 202
DB 481 CTCGGCGGGTAAACGGCAACCCCTACCCGCTTGCCTAAGCTCGGCAATACGAGGGGCC 540
QY 203 ValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIle 222
DB 541 GTCAACCCCGATACCCGCCACCGCGATTTGAACGTCGTCGTGGACACGCGGCCCATC 600
QY 223 AlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSerGly 242
DB 601 GCCTTCGGCGACTTTCGAAATCACCGGCACACAGGTTACCCCGCAACAAATTCGTCTCGGC 660
QY 243 LeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGln 262
DB 661 CTTGCGCGTTTCAGCGCGGTATGCCGTACGACCTCGACCTGCTCGACTTCCAACAG 720
QY 263 AlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeu 282
DB 721 CGGCTCGAACAAACGGGCAATTTCCCGCGCGTCCGTACAGCCGACTTCGACCGCCCT 780
QY 283 GlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGlu 302
DB 781 CAAGGGCAGCGGCTCCCGCTCAAAGTCAAGCGTAAACGAGGTCAAAACGCCCAAACTCGAA 840
QY 303 ThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 322
DB 841 ACCGGCATCCGCTCGATTCGGAATACGGTTTGGCGGCAAAATCGGCTACGACTATTAC 900
QY 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr 342
DB 901 AAGCTCTTCAACAAGGCTATATCGGTTCCGTTCGGGATATGGACAAATACGAAC 960
QY 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
DB 961 ACCTTTCGCCCGCGCATCAGCCAGCGCGCACTATCGGGCAACTACTCGGACAGCAAC 1020
QY 363 ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyIle 382
DB 1021 GTTTCTCTACAAACGTTTCGACCAACCCCAAACTCGGCAAAACCGCGCTTCTCCGCGCGTC 1080
QY 383 TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu 402
DB 1081 TGGTATGTCGGGACCGCGCGGCATCATGCCAGGCTGGGCGCGAATTTCTTCGAGAA 1140
QY 403 GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr 422
DB 1141 GCGCGGAAATCCCGCGCTCGGCTGTCGATTTGGGCAACACGCCACGCGCATGTGCTC 1200
QY 423 AlaSerTrpLysArgGlnLeuLeuAsnValLleuHisProGluAsnGlyHisTyrLeu 442
DB 1201 GCCTCTTGGAAACCGCAGCTCTCAACACCTGCTGCATCCCGAAACGCCCATTACCTC 1260
QY 443 AspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThr 462
DB 1261 GACGGGAAATCCGTCAGACTTTGGGCACATTCCTGCTCCACCGCGCTGATCCGAC 1320
QY 463 SerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIle 482
DB 1321 TCTGCCCGTGCAGTTATTTCTTACGCGCGCGGTTACGAACTCGACAGCATCGGACTTGC 1380
QY 483 ArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet 502
DB 1381 CGCGGCAACAGCGGGTTACACCGCTTCGCCGCGACAAATCCGACGCTTCCTTCAGGGCTGATG 1440
QY 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
DB 1441 TTCCGCGAGCGCGCGCTCTTCGCTGCGCGGTTACGAACTCGACAGCATCGGACTTGC 1500
QY 523 GlyProAsnGlySerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGln 542
DB 1501 GGCCCCGAACGGATCGGCTCTGCCGGAACGCGCCCTCTCTGGTGGGCGAGCCCTGGAATACCAA 1560
QY 543 LeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 562

```


|||||
Db 1561 CTGCGTTTACGGCACCCCTTTCGGCGCGGTCTTCCACGATATGGCGATGGCGCGCC 1620
QY 563 AanPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerPro 582
|||||
Db 1621 AATTTTCAAACGATATGAAGCTGAACACACGGCTTCGGGACTGGGGGTGGCTGGTTTCAGCCCG 1680
QY 583 LeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHis 602
|||||
Db 1681 CTTGGCGCGTTTCTTCCTCGACATCGCTACGGGACAGCGATGAAGAAATCCGCTGGCAC 1740
QY 603 IleSerLeuGlyThrArgPhe 609
|||||
Db 1741 ATCAGCTTGGGAACACGCTTC 1761
RESULT 8
AAZ53622
ID AAZ53622 standard; DNA; 1569 BP.
XX AC AAZ53622;
XX XX
DT 21-MAR-2000 (first entry)
XX Neisseria gonorrhoeae ORF 286 partial DNA sequence SEQ ID NO:1193.
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX XX
OS Neisseria gonorrhoeae.
XX XX
PN W09957280-A2.
XX XX
PD 11-NOV-1999.
XX XX
PF 30-APR-1999; 99WO-US09346.
XX XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin R, Venter JC;
XX XX
DR WPI; 2000-062150/05.
XX XX
XX P-PSDB; AAY74860.
XX XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX XX
PS Claim 7: Page 666; 1453pp; English.
XX XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.
XX Sequence 1569 BP; 388 A; 522 C; 393 G; 266 T; 0 other;
Alignment Scores:
Pred. No.: 3.86e-237 Length: 1569
Score: 2653.00 Matches: 512
Percent Similarity: 99.42% Conservative: 2
Best Local Similarity: 99.03% Mismatches: 2
Query Match: 83.61% Indels: 1
DB: 21 Gaps: 0
US-09-857-669-2 (1-609) x AAZ53622 (1-1569)
QY 1 MetMetIleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyr 20
Db 19 ATGATGATCAAAACGACCGCCCTCTCTCGCGCTTATTTTCTTCGCGACGATAC 78
QY 21 AlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLys 40
Db 79 GCGCTGCGCGCGACTTTCGAAACAAGCGCGGGTTCGCATTGTTCAAAAGCAAA 138
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 139 ACCCCGACACCGATCAGTCNAATTAACCCNAATTCCTCCGTCGATCCGACGAG 198
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 199 GACAGTGAATCAAGATATGGTCGAAGAACACTGCGCGTCATCAGCAGCAGCAGGAA 258
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 259 GAGGTTTGGATAAGGAACACACGGGATTCCTTCCGAGAAGACCCGCAACGCTTAA 318
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly 120
Db 319 ACAATGCTCCGACGCAAGGCTATTTTTCAGCAGCAAGGTCAGCCTGACGAAAACACGGA 378
QY 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 379 GCTTATACGGTGCACATCACACCGCGCGCCGCGACCAAAATCGCAACGCTCGCGCC 438
QY 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 439 ATCCTCGCGGACATCTTTCAGACGGCAACTGCGCGAATACCTACCGCAACGCGTGGAA 498
QY 161 AsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThr 180
Db 499 AACTGGCAGCAGCGGTAGGCAGCGATTCGATCAGCAGCAGATTGGGAAAACGAAACT 558
QY 181 SerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArg 200
Db 559 TCCGTCTCGCGCGGTAAACGCGCAAGGCTACCGCTTCGCAAGTCGCGCAACACCGCG 618
QY 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArg 220
Db 619 GCGGCGGTCAACCCCGATACCCGCCGATTTGAACGCTTCGTGGACAGCGCGCG 678
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 679 CCATTTGCTTCGGCGGACTTTGAAATCACCGGCACACAGCGTTTACCCCGCAACAAACCGTC 738
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 739 TCCGCGCTGGCGGCTTCCACCGCGGCACGCCCTACGACCTCGACCTGCTCGACTTC 798
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 799 CAACAGGCGCTCGAACAACAGCGGCATTATTCGGCGCGCTCCGTACAAAGCCGACTTCGAC 858
QY 281 ArgLeu-GlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 859 CGCCTCCCAAGGGGACCGCGTCCCGCTCAAGTCAGCGTAACCGAGGTCAACCGCACAA 918

```
QY 300 sLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyrAs 320
|||||
Db 919 ACTCGAACCAGCATCCCGCTCGATTCGGAATACGGTTTGGGCGGCAAAATCGCCTACGA 978
QY 320 pTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
|||||
Db 979 CTATTACAACCTCTTCAACAAGGCTATATCGGCTCGTCTGGATATGCACAATA 1038
QY 340 rGluThrThrLeuAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpTh 360
|||||
Db 1039 CGAAACACGCTTGGCGCGCATCAGCCAGCGCGCAACTATCGGGGCAAACTACTGGAC 1098
QY 360 rSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerG1 380
|||||
Db 1099 AAGCAACGTTTCTTACAACCGTTCGACCACCAAAACCTCGAAAAACGCGCTTCTCCGG 1158
QY 380 yGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLe 400
|||||
Db 1159 CGGCATCTGGTATGTGCGGACCGCGCGGCATCGATGCCAGGCTGGGGCGGAATTTCT 1218
QY 400 uAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMe 420
|||||
Db 1219 CGCAGAAGCGGAAATCCCGCTCGGATGTCGATTGGGCAACAGCCACGCGCAT 1278
QY 420 tLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHi 440
|||||
Db 1279 GCTGACCGCTCTTGGAAACGCGAGCTGCTCAACAGCGTCTGCACCCCGAAACGGCCA 1338
QY 440 sTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuI1 460
|||||
Db 1339 TTACCTCGACGGCAAAATCGGACGACTTTGGGCACATTCCTGCTCCACCGCGCTAAT 1398
QY 460 eArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPh 480
|||||
Db 1399 CGCACCTCTGCGCGCGAGGTATTTCTTACGCGCGGAAACAAACAAACTCGGCACGTT 1458
QY 480 eIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAlaAspValProSerG1 500
|||||
Db 1459 CATCATACGCGACACAGCGGTTACACGGTTGACAGCGACAATGCCGATGTCCTCCGCG 1518
QY 500 yLeuMetPheArgSerGlyGlyAlaLeuSerValArgGlyTyrGluLeu 516
|||||
Db 1519 GCTGATGTTCCGACGCGCGCGGTCTTCCGTCGCGGTTACGAACCTT 1567

RESULT 9
AAF61044
ID AAF61044 standard; DNA; 1374 BP.
XX
AC AAF61044;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF06604.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (CUFA-) QUTAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKF2-) DKF2 DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDI2INISCHE HOCHSCHULE HANNOVER.
```

```
XX WPI: 2001-192469/20.
XX
PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria.
XX
PS Claim 1a: Page 94; 158pp; German.
XX
CC This invention describes novel DNA sequences (I) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (1); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (1) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (1) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
CC carrying one or more (1). (1), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (1),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 1374 BP; 264 A; 428 C; 429 G; 251 T; 2 other;

Alignment Scores:
Pred. No.: 4.18e-26 Length: 1374
Score: 384.50 Matches: 115
Percent Similarity: 42.52% Conservative: 81
Best Local Similarity: 24.95% Mismatches: 224
Query Match: 12.12% Indels: 41
DB: 22 Gaps: 11

US-09-857-669-2 (1-609) x AAF61044 (1-1374)
QY 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaVal 186
|||
Db 43 GGTGAACAACTCAATCACGGCCATTACKAGGATGCAAGCGTTGATCCAGAACACGCGC 102
QY 187 ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
|||||
Db 103 TCGCGCTATGGCTTCTTCAGTGGCGCTTCAGTAGCAGCGCGCTCGACCGCGCA 162
QY 207 ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAsp 226
|||||
Db 163 GCGGGTGGCGGATATCGAACTGGTCTACAGAGTGGCCGCGTTATCGCCTGGCGGC 222
QY 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
|||
Db 223 GTCACCTTCGGTGGCGCACACCCCTGGACGAGGACCTGCTGCAGCGCATGTGTGCTTC 282
QY 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln 266
|||||
Db 283 AAGCGGGTACCCCTTACGACTCGGAACTGGTCCGACAGCTGAACACGACTCGCAATCG 342
QY 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArg 286
|||||
Db 343 AGCGGCTATTTCGAAGCGGTGGCGGTGGATGCGGCGGCCCACTGCTGCTGTGTCGGAAGA 402
QY 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyTleArg 306
|||||
Db 403 GTCCCGGTGGATGTTCACTGCAAAACCGTAAACACGACCATGGCGCTTGGCCTGGCGC 462
```

```
QY 307 LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrAsnLeuPheAsn 326
Db 463 TTCGACGCGAGTCGGCGCGCGCGCAAGTGGAGCCGCACTGGGTCAACCCA 522
QY 327 LysGlyTyrIleGlySerValTyrAspMetAspLys----- 339
Db 523 CAAGGCCAC-----ACGTATGCTGGGAACCGAACTCTCGGCCCGCCGCAACGTC 576
QY 340 -----TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 577 GGCCTGTGGTATGATCCCTCGACCGCGCGCTGACGACAAGTTCGCTTTCGCGCG 636
QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 637 GGTACCAAGACGAGGACTT-----GCCGGCACCCGACACGCTCAGCAAGCTG 684
QY 377 AlaPheSerGlyGlyIleTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 685 TTGACGGTCCGCCCGCGAGTGGACAGCAAGCTGCCAGTGGCTGGCAGCGGTGATTTCG 744
QY 397 AlaGluPheLeuAlaGlyArgGlyIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 745 CTCAGTAGTACCAACCGAAGATATCCCTG---GGTGCAGACTCCGCTTTCGCAACCTG 801
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 802 CTGATGCGGGCGGTACGCTTTCCTTCCTGCGCAGT-----GACAACCGTATCGATCCG 855
QY 437 GluAsnGlyHisTyrLeu-----AspGlyLysIleGlyThr 448
Db 856 CACAACGGCTATCCCTGCAGTTCAGTGTTCAGGTGCCAAGGAGGCTGTCGCGAC 915
QY 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 916 ACC-----AACCTGCTGCACGGCAGCTATTCCTCAAGGCGCTGACCACACTCGGCCAC 969
QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla----- 486
Db 970 -----AACCCAGCGCTTCTCGGACGCGTACACTTTTGGTGGCAGTGCACCAAT 1017
QY 487 GlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGly 506
Db 1018 GCTTC-----AGAACACATTCGCCCTCGCTCGCTTCCTTCGCGGT 1062
QY 507 GlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGly 526
Db 1063 GCGACACGAGTGTGCGGTTACGACTACACACCTCTCGCCGAAGAACAGCAGCGT 1122
QY 527 SerValLeuProGluArgAlaLeuValGlySerLeuGlyThrGlnLeuProPheThr 546
Db 1123 GACCTATCGCGGGCGCTACTTGTGGGAGGAGGTGTGCGATACCACTATTCGCTGACC 1182
QY 547 ArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArg 566
Db 1183 GAAAAATGGCGGTCGCGACTGCTCGACCAAGCAACTCGTTCACACGACCTGGAGCTG 1242
QY 567 MetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe 586
Db 1243 CCCACCTCAAGACGGGTGCTGGTGTGGCTGGGTATCGCCAGTGGCGGCGGTG 1302
QY 587 SerPheAspIleAlaTyrGlyHisSerAspLys---LysIleArgTrpHisIleSerLeu 605
Db 1303 CGCCTGGACCTGGCAAGCGCTGGATGACGAAGGGGCGATTGCGCTGCACTTTTCCATG 1362
QY 606 Gly 606
Db 1363 GGG 1365
RESULT 10
ID AAA50269
XX AAA50269 standard; DNA; 1737 BP.
```

```
AC AAA50269;
XX 07-NOV-2000 (first entry)
DE Haemophilus influenza strain Rd KW20 BASB067 gene.
KW BASB067 gene; outer membrane protein; vaccine; antibiotic;
KW antibacterial; screening; infection; diagnosis; therapy; ds.
XX Haemophilus influenzae.
OS
FH Key Location/Qualifiers
FT CDS 1..1737
FT sig_peptide /*tag= a
FT FT 1..66
FT mat_peptide /*tag= b
FT FT 67..1734
FT FT /*tag= c
XX WO200047737-A1.
XX 17-AUG-2000.
XX 04-FEB-2000; 2000WO-EP00887.
XX 09-FEB-1999; 99GB-0002880.
XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Ruelle J, Thonnard J;
XX WPI; 2000-515059/46.
XX P-PSDB; AAY95820.
XX BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
XX used for diagnosing and treating H. influenzae infections -
XX Claim 3; Page 77-78; 87pp; English.
XX The present sequence is that of the coding region of the BASB067
XX gene of Haemophilus influenzae strain Rd KW20. It encodes a
XX 578-amino acid surface expressed protein (see AAY95820) that is
XX recognised by the immune system. The invention relates to
XX recombinant materials and methods for the production of BASB067
XX polypeptides and polynucleotides, for use especially in therapeutic
XX and prophylactic vaccines. It also relates to methods for using
XX such polypeptides and polynucleotides in the prevention and
XX treatment of microbial diseases, in diagnostic assays for detecting
XX diseases associated with microbial infections, and assays for
XX detecting expression or activity of BASB067 polypeptides or
XX polynucleotides. A polynucleotide having at least 85% identity
XX to the present sequence can be used in the recombinant production
XX of BASB067 immunogenic polypeptides in transformed host cells, and
XX in vaccine compositions.
XX Sequence 1737 BP; 517 A; 309 C; 380 G; 531 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 5,46e-25 Length: 1737
Score: 374.00 Matches: 153
Percent Similarity: 36.88% Conservative: 93
Best Local Similarity: 22.94% Mismatches: 263
Query Match: 11.79% Indels: 158
DB: Gaps: 18
US-09-857-669-2 (1-609) x AAA50269 (1-1737)
QY 3 IleLysProThrAlaLeuLeuProAlaLeuPhePheProHisAlaTyrAlaPro 22
Db 16 CTAAACTCAGCGCACTTTTATTA---GCATTAGTGTCTTCCCGCAACAA 72
QY 23 AlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerPro 42
|||||
|||||
```

```

Db 73  ACCGTTGATATTCAAGGCAATTCGTGGTTTCGTCGTGCGTAATACAGAT--- 129
QY 43  AspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSer 62
Db 130 -----CTCAATGTTAAATTTGATTAATAAA 153
QY 63  GluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluVal 82
Db 154  GAAGAAATGGATGGCTCCGGAACGCTATCAACATTTGGTAACCAAGCC----- 201
QY 83  LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
Db 202 -----GTGGATCGTGGT 213
QY 103  LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAsp----- 119
Db 214  TTGCGGTGTTGTTGTTATTAATCAATCTCCGTGGGTTTTGAACGAAACACAGCGTCAAGCC 273
QY 120 -----GlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137
Db 274  AAMCGCATTTATGATGCTCATGTTACACGAGCGGAGCCAAACAAAATTCGGGGACT 333
QY 138  GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
Db 334  GATGTGCAAAATTCAGGGGAAGCCGCAAGATCAAAATTTT-----AAT 378
QY 158  AlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
Db 379  GCATTCGGTAAAAACTTACCGAAGATGGCGTTTGTGAGCACCACAACTTACGATGAT 438
QY 178  SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 439  TACAAAACAGCATTCACGTTTAGCATTAAATCGTGGGTATTTTCATGGGAACTTTAAA 498
QY 198  AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
Db 499  ATTTACAGTTTAAAGATCAGCCCTGAACCCATCAAGCATGGTGGCGAATGTTATTTGAT 558
QY 218  SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
Db 559  AGTGGTGTCCGTATCATATATGCAATATTTACTTTTAGCCATTCACAA---ATTCGGGAT 615
QY 238  GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 257
Db 616  GATTACTTAATTAATTTCTTAACATCAATCTGGTGATCCGTAATTAATGAATAATTG 675
QY 258  LeuAspPheGlnGlnAlaLeuGluAlaAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
Db 676  TCGGATTTAACCAGCGATTTCCATCTTCAAAATTTGGTTAGCTCAGTATTAGTTCAGCT 735
QY 278  AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
Db 736  AATGTTAATCAT---AAAGCAAAACTGTGGATGGAGATTAATCTTTATCCACGTAHA 792
QY 298  ArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle 317
Db 793  AAAAATGCGATGGAACCTCGGTGGGTGTTCTACTGATGGCGCGTTTCACGCAACAAATA 852
QY 318  AlaTyr----- 319
Db 853  GGCTGGACAAAACCTTGGATTAAAGCCGTGGACATAGTTTGGTTCAAAATCTTTATCTC 912
QY 319 ----- 319
Db 913  TCTGCACCAAAACAACTTAGAGGCAACTTATCGAATGCCACTGCTTAAAAATCCATTA 972
QY 320  AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 973  AATTAATGATGATTTTCCGCTCGGTGGAGGGGAAAAAGAAATGATACCAATACG 1032
QY 340  TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
Db 1033  AGAGTGCTTACGTTGTCAGCG-----TTACGTTATTGG 1065

```

```

QY 360  ThrSer-----AsnValSerTyrAsnArgSerThr 369
Db 1066  AATAATGGCGCATGGTTGGCAATATTTTGGCGGACTTCGTATGCATACGACAGTTTTTACA 1125
QY 370  ThrGlnAsnLeuGluLysArgAlaPhe-----SerGlyGlyIleTyrTrpVal 385
Db 1126  CAAGCGGATATCATCATGATAAAACCTTACTCTTTATCCAACTGTTGGATTACTCGTACT 1185
QY 386  ArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLys 405
Db 1186  CGATTACGTTGGTGGTTCCTTTGCCACTTTGGCC-----GATGTGCCAAAAA 1230
QY 406  IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrp 425
Db 1231  ATT-----ACTTTTGAATTAGC----- 1248
QY 426  LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1249  AACGA----- 1254
QY 446  IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1255  -----ATTGGCTATCACAATCTTTTATAAAAGTGCAAGCATCT 1296
QY 466  AlaGlyTyrPhePheThr-----ProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
Db 1297  ACGCGCTGGTTCGTACTTATGCAGAAATCATCGTGC-----GTTGCTCGTCT 1347
QY 485  GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 1348  GAAATCGCGTATTATCATACAAAAGGTATTTCAAAAAATTCGGCTACACTGCGTTCTTT 1407
QY 505  SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
Db 1408  GCTGSGGCGATCGTAGTGTGCGCGGTATACGGCTATAAAAAATTCGGCTAAAAATAGA 1467
QY 525  AsnGlySerValProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuPro 544
Db 1468  AATGGAATAATGCTGGTGGTTCGCGATTGCTTACCACCTCTTTAGAAATCAATATCAA 1527
QY 545  PheThrArgThrLeuSerGlyValAlaValPheHisAspMetGlyAspAlaAlaAsnPhe 564
Db 1528  GTCTATCCAAATTTGGTGGCGGCAACTTTTCAGATAGTCGATTAGCTGCCGATAATTAC 1587
QY 565  LysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 1588  ACAGCAAAAGAGCTCGCTTATGGCAGCGGCTGCTGCTGGCTGGCAGTGGT 1647
QY 585  ProPheSerPheAspIleAlaTyrClyHisSerAsp-----LysLysIleArg 600
Db 1648  GCGATTAAATTTGATTTGCCACACCCCATTCGTGTATAAGATAACAGCAAAATATTCAA 1707
QY 601  TrpHisIleSerLeuGlyThr 607
Db 1708  TTTTACATCGGACTTGGTACA 1728
RESULT 11
AAA50270
ID AAA50270 standard; DNA; 1731 bp.
XX
AC AAA50270;
XX
DT 07-NOV-2000 (first entry)
XX
DE Haemophilus influenza non-typeable strain 289 BASB067 gene.
XX
KW BASB067 gene; outer membrane protein; vaccine; antibiotic;
XX
OS antibacterial; screening; infection; diagnosis; therapy; ds.
XX
FH Haemophilus influenzae.
Key Location/Qualifiers

```

```

FT CDS 1..1731
FT /*tag= a
FT sig_peptide 1..60
FT /*tag= b
FT mat_peptide 61..1728
FT /*tag= c
FT
FT PN W0200047737-A1.
XX
XX PD 17-AUG-2000.
XX
XX PF 04-FEB-2000; 2000WO-EP00887.
XX
XX PR 09-FEB-1999; 99GB-0002880.
XX
XX PA (SMIRK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX PI Ruelle J, Thonnard J;
XX
XX DR WPI: 2000-515059/46.
XX
XX DR P-PSDB; AAY95821.
XX
XX BASB067 polypeptide and polynucleotide from Haemophilus influenzae are
XX used for diagnosing and treating H. influenzae infections -
XX
XX PT Claim 3; Page 80; 87pp; English.
XX
XX CC The present sequence is that of the coding region of the BASB067
XX gene of Haemophilus influenzae non-typeable (Nthi) strain 289. The
XX DNA was isolated by PCR using primers (see AAY50271-72) specific for
XX the BASB067 gene. It encodes a 576-amino acid surface exposed
XX protein (see AAY95821) that is recognised by the immune system. The
XX invention relates to recombinant materials and methods for the
XX production of BASB067 polypeptides and polynucleotides, for use
XX especially in therapeutic and prophylactic vaccines. It also
XX relates to methods for using such polypeptides and polynucleotides
XX in the prevention and treatment of microbial diseases, in diagnostic
XX assays for detecting diseases associated with microbial infections,
XX and assays for detecting expression or activity of BASB067
XX polypeptides or polynucleotides. A polynucleotide having at least
XX 85% identity to the mature portion of the present sequence
XX (nucleotides 61-1731) can be used in the recombinant production
XX of BASB067 immunogenic polypeptides in transformed host cells,
XX and in vaccine compositions.
XX
XX SQ Sequence 1731 BP; 509 A; 322 C; 386 G; 514 T; 0 other;

Alignment Scores:
Pred. No.: 2,86e-23 Length: 1731
Score: 355.50 Matches: 138
Percent Similarity: 38.01% Conservative: 87
Best Local Similarity: 23.31% Mismatches: 238
Query Match: 11.20% Indels: 129
DB: 21 Gaps: 15

US-09-857-669-2 (1-609) x AAA50270 (1-1731)
QY 79 GlnGluValAlaLeuAspLysGlnThrGlyPheLeuAlaGluAlaProAspAsn 98
Db 145 AAGAAGAAGTGGCGGTGTAACGCTATCAGCATTTGGTAACCAAGCCGGGATCCT 204
QY 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLys 118
Db 205 -----GGTTGGCTGCTGTTGGTTATTATCAATCTCCGTCGCTTTTGAACGAAA 255
QY 119 Asp-----GlyAlaTyrThrValHisIleThrProGlyProArgThrLys 133
Db 256 CACCGTCAAGGCAACGGCGATTATTGATTGCTCATGTTACACGCGAGGCCAACAAA 315
QY 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAlaGlu 153
Db 316 ATTGCGGGCACTGATGCAAAATTGAGGGGGAAGCCGCACAAAGATGAATTTT----- 369

QY 154 TyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAsp 173
Db 370 -----GATCGCTACTAGTAAAACTTGCACAAAGAGCGCTTTGGTTGAACACCAA 420
QY 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeu 193
Db 421 ACTTACGATGATTACAAAACAGCGATTTCACGCTTGCATTAATAATCGTGGTATTGTGAT 480
QY 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
Db 481 GGAACCTTTAAATTTTCACGTTTACAAATGACCCCTGAAACCCATCAAGCATGGTGGCGA 540
QY 214 ValValAlaAspSerGlyArgProIleAlaPheGlyPheGluIleThrGlyThrGln 233
Db 541 ATGTTATTGTAGTGGTGTCCGTTATCATTAATGCAATATTACTTTTATGCCATTCACAA 600
QY 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
Db 601 ---ATCCGTGACGATTATCTCAATAATATCTTAACATCAAACTGGCGATCCATATTATTA 557
QY 254 LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 658 ATGAATAATTTCTCGGATTAAACACGCGATTTTTCATCTTCAAAATGGTTAGCTCAGTA 717
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
Db 718 TTAGTTTCACGCTTAATGTTAATCAT---AAAAGCAAAACTGTGGATGTGGAGATTATCTTT 774
QY 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
Db 775 TATCCACGTAAAAAAAATGCGATGGAACCTCGTGTGGCTTGTGCTACTGATGCGCGCTT 834
QY 314 GlyGlyLysIleAlaTyr----- 319
Db 835 CACGGACAATTTGGCTGGACAAAACCTTGGATTAAACCGCTGGACATAGTTTGCCTTCA 894
QY 319 ----- 319
Db 895 AATCTTTTATCTCTCTGCACAAAACAACTAGAGGCAACTTATCGAATGCCACTGCTT 954
QY 320 -----AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
Db 955 AAAAAATCCATTAATATTATTACTATGATTTTGCCTGCTGGGAGGGGAAAAAGAGAAC 1014
QY 336 AspMetAspLysTyrGluThrThrLeuAlaGlyIleSerGlnProArgAsnTyrArg 355
Db 1015 GATACCAATACGAGAGCTGCTTACGTTGTACGCG----- 1047
QY 356 GlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLys 375
Db 1048 TTACGTTATTGG-----AATAATGCGCATGGT 1074
QY 376 ArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
Db 1075 TGGCAATATTTTGGCGGACTT-----CGTACGCGATACGACAGTTTATACA 1119
QY 396 GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsn 415
Db 1120 CAAGCGGATATC----- 1131
QY 416 SerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHis 435
Db 1132 ACTGATAAAACCTTACTT-----CTTTAT 1155
QY 436 ProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThr----- 452
Db 1156 CCAACTGTTGGATTACTCGCACTCGATACGTGGTGGTCTCTTGGCATCTGGGGCGAT 1215
QY 453 -----PheLeuSerSerThrAlaLeuIle 460
Db 1216 GTCAAAAAAATTAATTTTGTATTAAAGCAAAACGAATTTGGCTATTCAGAATCTTCTTTATA 1275
QY 461 ArgThrSerAlaArgAlaGlyTyrPhePheThr---ProGluAsnLysLysLeuGlyThr 479

```

```

Db 1276 AAAGTCAAGCATCTAGCGGTGGGTTCTGTTACTTATGCAGAAATCATCTGATC----- 1329
Qy 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaIleArgAspAsnAlaAspValProSer 499
Db 1330 ---GTTGCTCGTCTGAAATCGGATTTTACATACAAAGAAGATATTGAAAAAATTCGCGCT 1386
Qy 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 1387 ACACCTCGTTCTTCTGCTGGCGCATGCTAGTGTGGCGGTTACGGCTATAAAAAAT 1446
Qy 520 GlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 1447 GCCTCTAAAAATAGAAATGAAATGTTGGCGGCTACCGCTTGTCTTACCATCTCTTA 1506
Qy 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 559
Db 1507 GAAATCAATATCAAGTTTATCCAAATGTGTGGCGGCACTTTTGCAGATAGTGATTA 1566
Qy 560 AlaAlaAlaAsnPhelysArgMetIysLeuIysHisGlySerGlyLeuGlyValArgTyr 579
Db 1567 GCTGCCGATAATTACACAGCAAAAGAGTGGTTATGGCACAGCGCTTGTGTGCGTTGG 1626
Qy 580 PheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAsp----- 596
Db 1627 GCATCCCGCAGTGGTGGCGGATTAATAATTCATATTCGCACACCCATTCGTGATNAAGATAAC 1686
Qy 597 ---LysIleIleArgTrpHisIleSerLeuGlyThr 607
Db 1687 AGCAAAAATTAATTTTACATCGGACTTGGTACC 1722

```

RESULT 12

AAAL1515

ID AAAL1515 standard; DNA; 2505 BP.

AC AAAL1515;

XX AAAL1515;

XX 21-AUG-2000 (first entry)

XX DNA encoding outer membrane protein (omp) 85.

DE Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
 KW meningococcal infection; protective immune response; vaccine; ss.
 XX Neisseria gonorrhoea.

XX Key Location/Qualifiers

FH RBS 49..53

FT /*tag= a

FT 5'UTR 1..59

FT /*tag= b

FT CDS 60..2438

FT /*tag= c

FT /product= "outer membrane protein 85"

FT sig_peptide 60..122

FT /*tag= d

FT 3'UTR 2439

FT /*tag= e

XX WO200023595-A1.

XX 27-APR-2000.

XX 22-OCT-1998; 98WO-US22352.

XX 22-OCT-1998; 98WO-US22352.

XX (UIMO-) UNIV MONTANA.

XX Judd RC, Manning SD;

XX WPI; 2000-339694/29.

XX P-PSDB; AAY84946.

DR

XX

PT New isolated outer membrane protein 85 of *Neisseria gonorrhoeae* and *N.*
 PT meningitidis useful for vaccine, therapeutic and diagnostic
 PT compositions for gonococcal or meningococcal infections -

XX Claim 6; Fig 2; 98pp; English.

XX

CC The present sequence encodes an outer membrane protein (omp) 85 of
 CC *Neisseria gonorrhoeae*. The omp polypeptides and polynucleotides are
 CC useful in compositions for use in the prevention, treatment and diagnosis
 CC of non-symptomatic gonococcal infection or meningococcal infection and
 CC symptomatic disease. They are also useful for the detection of
 CC hybridisation complexes. Antigens and antibodies specific omp proteins
 CC also provide diagnostic, therapeutic and prophylactic compositions for
 CC the treatment or prevention of the infections described above. The
 CC antibodies are useful for inducing a protective immune response in
 CC humans or animals with *N. gonorrhoeae*, *N. meningitidis*, or other
 CC *Neisseria* species. The proteins, antibodies and polynucleotide
 CC sequences of the present invention may also be used in the screening
 CC and development of chemical compounds such as drugs or vaccines.

XX

SQ Sequence 2505 BP; 669 A; 787 C; 621 G; 428 T; 0 other;

Alignment Scores:

Pred. No.:	3-29e-14	Length:	2505
Score:	260.50	Matches:	145
Percent Similarity:	36.53%	Conservative:	86
Best Local Similarity:	22.94%	Mismatches:	245
Query Match:	8.21%	Indels:	157
DB:	21	Gaps:	28

US-09-857-669-2 (1-609) x AAAL1515 (1-2505)

Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys-----ThrMet 102

Db 685 GAAGCGACCGGTTCGACCGCCAGAAATTCGCCCAAGACATGGAAAAAGTAACGACTTCT 744

Qy 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117

Db 745 ACCAGAACAC-GGCTACTTCGATTCGATATCCGACATCCCAACCAACGAA 803

Qy 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135

Db 804 GACAAAACCGCAGCAGCACCATCAAAATCACCGTCACGAAGCGCGAGCTTTCCGCTGGGC 863

Qy 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155

Db 864 AAAGTGTGCAATT-----GAAGCGCAGACCAACGAAGTCCGC 899

Qy 156 ArgAsnAlaLeuGlu-----AsnTrpGlnProValGly 167

Db 900 AAGCGCGAATCGAAAACTGCTGACCATGAAGCCCGCAATGGTACGAA----- 950

Qy 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187

Db 951 -----GCCCAGCAGATGACCGCGCTTTTGGGTGAGATTTCAG 986

Qy 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203

Db 987 AACCGCATGGGTCGGCAGGTACGCATACAGCGAATC---AGCTGACCGCGCTGCCG 1043

Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAla 223

Db 1044 AACCGCGAACCACCAACCGCTCATTTCTGTCGACATCGAACCGCGCAGAAAAATCTAC 1103

Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243

Db 1104 GTCAACGAAATCCACATACCGCGCAACCAACAAACCCGCGACGAGTCTGCGCCCGAA 1163

Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263

Db 1164 TTGCGCCAAATGGAATCCGCGCTTACGACACCTCCAAAGCTGCAACGCTCCAAAGCGC 1223

```
QY 264 LeuGluInAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheArgLeuGln 283
Db 1224 GTCCAGCTTTGGGCTACTCTCGAC-----AACGTACAGTTTGATGCGCTCCCGCTGCC 1277
QY 284 Gly-----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 1278 GGTACGCCGCCCAAGTCGATTTGAACATGAGCTGACCGAAGCTTTCACCGGCTCGCTC 1337
QY 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
Db 1338 GACTTGACGGCGGCTGGTTCAGGATCGGCTTGTCATGTCGCGCGGCTATCGGAG 1397
QY 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValTyrAspMetAspLysTyrGlu 341
Db 1398 GACAAACCTGTTCCGTACGGC-----AAGTCGGCGCGCTCGCGCCCTCGCGAAGCAAA 1451
QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrThrPheSer 361
Db 1452 ACCACGCTCAAGGCTCGCTGTCG-----TTTACCGACCCGTACTTTCACGGCA 1499
QY 362 -----AsnValSerTyrAsn-----ArgSer 368
Db 1500 GACGGGCTCAGCCTGGCTAGCATATTACGGAAGAGCTTTCACCGCGCGCAAGCATCG 1559
QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
Db 1560 ACCACGCTCAACATATATAAACACACCGCGCGCGGCGGTAGAGATGGGTATCCCC 1619
QY 382 ---IleTyrThrValArgAspArgAlaGlyIleAspAla----- 393
Db 1620 GTTACCGAATACGACCGCTCAATTTCCGGCTGGCGCGGCAACACCTGACCGTCAACACC 1679
QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGlyArgLysIleProGly 408
Db 1680 TACACAAAGCACCCAAACGCTATCCGACTTTATCAACATATACGGCAAAACCGACGGC 1739
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425
Db 1740 GCAGAC-----GGCAGCTTCAAAAGGCTGTGTACAAAGGCACCTGTCGGCTGG 1787
QY 426 LysArgGlnLeuLeuAsnAsnValIleHisProGluAsnGlyHisTyrLeuAspGlyLys 445
Db 1788 GGGCGCAACAGACGACGCGCTTATGGCGGACGCGGCGC---TACCTGACCGGC--- 1841
QY 446 IleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArg 465
Db 1842 -----GTAATGCGCGAAATCGCCCTGCGCGCGCAAGCAAA 1874
QY 466 AlaGlyTyrPhePheThrProGluAsnLys-----LysLeuGly 478
Db 1875 CTGCATACTACTCGCCGCCACCAACCAACCTGCTTTCCTTAAAGCAAAACCTTC 1934
QY 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAla---AspVal 497
Db 1935 ACGCTGATGCTCGCGCGGAAAGTCGGCATTCGCGGCGGTACGCGCAACCAAGAAATC 1994
QY 498 ProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAsp 517
Db 1995 CCTTCTTTGAAAATCTTACGGCGGCGGCTGGTTCGGTTCGGCGGTACGAAAGCGGC 2054
QY 518 SerIleGly-----LeuAlaGlyProAsnGly 526
Db 2055 ACGCTCGGCGCCAAAGTGTATGACGAATACGGCGGCAAAATACGTCACGGCGGCAAC--- 2111
QY 527 SerValLeuProGluArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2112 -----AAAAGCCAAAGCTTCGCGCGAGCTGCTTCCCGATGCGCGGTGGG 2159
QY 545 ---PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2160 AAAGACGACGACCGCTCCGCTGAGCTGTTTCCCGACGCGACGCGGTGTGGGACGGC 2219
QY 560 -----AlaAlaAlaAsnPhelysArgMet----- 567
```

```
Db 2220 AGAACCTATACCGCGCGGAAGCGGTAAACAACAATCGGTTTACTCGGAAAACCGCAT 2279
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
Db 2280 AAATCCACCTTTACCAACGAATTCGCTATTCCGCGCGCGCGGCTTACCTGGCTCTCG 2339
QY 582 ProLeuAlaProPheSerPheAspIleAlaTyr-----ClyHisSerAspLys 597
Db 2340 CCTTTGGCGCGATGAATTCATCTACGCTACCGCTGAGGAAACCGGAAACGCA 2399
QY 598 LysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2400 ATCCAACGCTTCCAATTCAGCTCGGCACGAGTTC 2435
RESULT 13
ARS07278
ID AAS07278 standard; DNA; 2379 BP.
XX
AC AAS07278;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein DNA.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2379
FT /tag= a
FT /product= "N. meningitidis serogroup A antigen"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..2376
FT /tag= c
FT /product= "Mature N. meningitidis serogroup A antigen"
XX
WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 95GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
XX (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
XX Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX WPI: 2001-381289/40.
XX P-PSDB; AAU03958.
XX
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
XX gonorrhoeae, useful in the manufacture of a medicament for treating and
XX preventing Neisserial bacteria infection -
XX Claim 6; Fig 2: 92pp; English.
XX
XX The sequence represents a DNA encoding a Neisseria meningitidis serogroup
XX A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx,
XX causing meningitis and, occasionally, septicaemia in the absence of a
XX meningitis. This antigenic protein is useful in the manufacture of a
XX medicament for treating or preventing infection due to Neisseria
XX bacteria, such as meningitis and septicaemia. It is also useful as a
XX diagnostic reagent for detecting the presence of Neisseria bacteria or
XX antibodies raised against Neisseria, and as a reagent for raising the
XX antibodies. The Neisserial nucleotide sequences can be expressed in a
XX variety of different expression systems, for example, mammalian cells,
```

25 ACCACGCTCATCCGCTCGCTGTCG
 1555 -----TTACCGACCCGTACTTCACGGCA

AC AAH42129;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE Nucleotide sequence of a *Neisseria gonorrhoeae* protein.
 XX
 KW Serogroup B protein; outer membrane protein; *Neisseria* infection;
 KW vaccine; ss.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2379
 FT /*tag= a
 FT /product= "Neisseria serogroup B protein"
 FT 1..63
 FT sig_peptide /*tag= b
 FT 64..2376
 FT mat_peptide /*tag= c
 XX
 PN WO200152885-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 17-JAN-2001; 2001WO-IB00166.
 XX
 PR 17-JAN-2000; 2000GB-0001067.
 PR 09-MAR-2000; 2000GB-0005699.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Pizza M, Rappuoli R, Giuliani M;
 DR WPI: 2001-451895/48.
 DR P-PSDB; AAB84745.
 XX
 PT Composition for treating or preventing infection to, detecting, or for
 PT raising antibodies against *Neisseria* bacteria, comprises an N.
 PT meningitidis serogroup B outer membrane preparation and an immunogenic
 PT component.
 XX
 PS Disclosure; Page 64-65; 83pp; English.
 XX
 CC The present sequence encodes a *Neisseria gonorrhoeae* protein. The
 CC protein is used to produce the compositions of the invention. The
 CC specification describes a composition, comprising a *Neisseria*
 CC meningitidis serogroup B outer membrane preparation and an immunogenic
 CC component. The immunogenic component is protein disclosed in WO99/57280,
 CC WO99/36544, WO99/24578, WO99/56791, WO97/28273, WO96/29412, WO95/03413,
 CC WO99/31132, WO99/58683, WO99/55873, and/or *N. meningitidis* protein PorA,
 CC TbpA, TbpE, PilC, OPA, or Omp85. The composition is used for making
 CC a medicament for treating or preventing infection due to *Neisseria*
 CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*
 CC bacteria or of antibodies raised against *Neisseria* bacteria; and/or
 CC a reagent which can raise antibodies against *Neisseria* bacteria. It may
 CC also be used as a vaccine.
 XX
 SQ Sequence 2379 BP; 636 A; 755 C; 596 G; 392 T; 0 other;

Alignment Scores:

Pred. No.:	7,23e-14	Length:	2379
Score:	256.50	Matches:	144
Percent Similarity:	36.39%	Conservative:	86
Best Local Similarity:	22.78%	Mismatches:	246
Query Match:	8.08%	Indels:	157
DB:	22	Gaps:	28

US-09-857-669-2 (1-609) x AAH42129 (1-2379)

QY 86 GluGlnThrGlyPheLeuAlaGluProAspAsnVallys-----ThrMet 102
 ||| ||||| ||| :||| |||
 Db 626 GAAGGACCGGTTCGACCGCCAGAAATTCCGCCAAGACATCGGAAGTAACCGACTTCT 685

QY 103 LeuArgSerLysGlyTyrPhe-----SerSerLysValSerLeuThrGlu 117
 |||::: ||||| |||
 Db 686 ACCAGAACAC-GGCTACTTTCGATTTCCGTATCTCGATCCAGCATCAACACCAACGAA 744
 QY 118 LysAspGlyAlaTyrThrValHisIleThr-----ProGlyProArgThrLysIleAla 135
 |||::: ||||| |||
 Db 745 GACAAACACCGACGACCAAAATCAGCTCCACGAAGCGGAGCTTTCGCTGGGCGC 804
 QY 136 AsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
 |||::: |||::: |||
 Db 805 AAAGTGTCTGATT-----GAAGGCGACACCAACGAATCCCTCC 840
 QY 156 ArgAsnAlaLeuGlu-----AsnTrpGlnGlnProValGly 167
 |||::: ||||| |||
 Db 841 AAGGCCGAACCTGGAAAAAATCTGACCATCAAGCCCGCAATGGTAAGAA 891
 QY 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
 |||::: ||||| |||
 Db 892 -----CGCCAGCAGATCCCGCTTTGGTGAGATTTCAG 927
 QY 188 ArgLys-----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
 |||::: ||||| |||
 Db 928 AACCGCATGGCTCGGCAGGCTACGCATACAGCGAAATC---AGCGTACAGCGCTGCCG 984
 QY 204 AsnProAspThrAlaThrValAspAsnValValAspSerGlyArgProIleAla 223
 |||::: ||||| |||
 Db 985 AAGCGCGAAGCAAAACCGCTGATTCGTCTGCATCGAACCGCGCGGAAATCTAC 1044
 QY 224 PheGlyAspPheGluIleThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
 |||::: ||||| |||
 Db 1045 GTCAACGAAATCCATCACCGCGCAACAAACCGCGACGAGCTGCGCGCGGAA 1104
 QY 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
 |||::: ||||| |||
 Db 1105 TTGCGCCAAATGGAATCGCGCCTTACGACCTCCAGCTCCACGCTCCACGCGCGC 1164
 QY 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
 |||::: ||||| |||
 Db 1165 GTCGAGCTTTTGGGTACTTCTCGAC-----AAGCTACAGTTTGTATCCCTCCCGTGGC 1218
 QY 284 Gly-----AspArgValProValLysValSerValThrGluValLysLeu 301
 |||::: ||||| |||
 Db 1219 GGTACGCGCGCAAGTCGATTGAACTAGCTGACCGAGCTCCACGCTCCACGCGCTCGCTC 1278
 QY 302 GluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGlyLysIleAlaTyrAspTyr 321
 |||::: ||||| |||
 Db 1279 GACTTTCGCGCGCGCTGGTTCAGGATACCGCTTGTCTCATGTCGCGCGCTATCGCAG 1338
 QY 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
 |||::: ||||| |||
 Db 1339 GACACCTGTTTCGGTACGGGC-----AAGTCGGCGCGCTCGCGCGCTCGCGAAGCAAA 1392
 QY 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
 |||::: ||||| |||
 Db 1393 ACCACGCTCAACGGCTCGCTCGC-----TTTACCAGCCCGTACTTTCACGGCA 1440
 QY 362 -----AsnValSerTyrAsn-----ArgSer 368
 |||::: ||||| |||
 Db 1441 GACGGGTTCAGCTCGGCTACGATATTACGGAAGCCCTTCGACCGCGCAAGCATCG 1500
 QY 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGly----- 381
 |||::: ||||| |||
 Db 1501 ACCAGCGTCAAAACAATATAAAACCAACCGCGCGCGCTAAGGATGGTATATCCCC 1560
 QY 382 ---IleTrpTyrValArgAspArgAlaGlyIleAspAla----- 393
 |||::: ||||| |||
 Db 1561 GTTACCGAATACGACCGCTCAATTTCCGGCTGGCGCGGAACACCTGACCGCTCAACACC 1620
 QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluArgLysIleProGly 408
 |||::: ||||| |||
 Db 1621 TACAAACAAAGCACCAACGCTATGCGGACTTTATCAGGAATACGGCAACCGACGCGC 1680
 QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSerTrp 425


```

QY 144 spileuSerAspGlyAsnLeuLaGluTyrTyrArgAsnAlaLeuGlu----- 160
Db 817 -----GAAGGCGACACCACCAAGTCCCAAGCCGCAACTGGAAACAACTGCTGA 865
QY 161 -----AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpG 176
Db 866 CCATGAAGCCCGCAATGGTACGAA-----C 892
QY 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrP 192
Db 893 GCCACAGATGACCGCGCTTTGGTGAGATTCCAGAACCGATGGCGCGGCGGCTACG 952
QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
Db 953 CATACAGCGAAATC---AGCGTACAGCGCTGCCAAACGCCGAAACCAACCGTCGATT 1009
QY 212 euAsnValValAspSerGlyArgProileAlaPheGlyAspPheGluLeuThrGlyT 232
Db 1010 TCGTCTGCATCGACCGCGCGGCGGAAATCTACGTCACCAAGATCCACATCACCGGCA 1069
QY 232 hrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
Db 1070 ACAACAAACCCGCGACGAGTCTGCGCGCGCAATTGGCCCAATGGAATCCGCGCCTT 1129
QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerG 272
Db 1130 AGGACCTCCAGCTGACAGCTCCAAAGAGCGCGTGGAGCTTTGGGCTACTTCGAC- 1188
QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
Db 1189 -----AACGTACAGTTGATCGCTGCGCGCTGCCGCGACACCCGACAAAGTCGATTGA 1243
QY 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
Db 1244 ACATGAGCTGACCGAAGCTTCCACCGGCTCGCTCGACTTGGCGGCGCTGGGTACAGG 1303
QY 310 luTyrGlyGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrI 330
Db 1304 ATACCGGCTGTGCTATGTCGCGAGCGTTTCCCAAGACAACCTGTTTCGCTACGGGC--- 1359
QY 330 leGlySerValValTrpAspMetAspLysTyrGluThrThrLeu----- 344
Db 1360 --AAGTCGCGCGCTGCGCGCTCAGCAAGCAACACCGCTCAACGGCTCGCTGCTGT 1417
QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnT 358
Db 1418 TTACCGACCGCTACTTACGCGACAGCGGCTGAGCTGGGCTACGATGTTTACGAAAG 1477
QY 358 yrTrp-----ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnL 373
Db 1478 CTTTGACCGCGGCAACATCGACAGCATCAACATATATAAACACCAACCGCGAGCGC 1537
QY 373 euGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspA 393
Db 1538 CAGGCATCGCATGAGCTGCTGTTACCGAATACGACCGCGTGAATTTTCGTTGGTGG 1597
QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
Db 1598 CAGACACCTGACCGCTCAACACCTACAAACAGCGCCCAACACTATGCGCGCTTATCA 1657
QY 401 laGluLysArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
Db 1658 AGAATACCGCAACACCGACGCGACAGAC-----GGCAGCTTCAAAGGCTGCTGT 1708
QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
Db 1709 ACANAGGTACCGCTCGGCTGGGGGCGCAACAAACCGGCGGCTTATGGCGCGCGCG 1768
QY 439 lyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
Db 1769 GC---TACCTGACGGC-----GTGAACGCGCAAA 1795

```

```

QY 459 euIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLys----- 475
Db 1796 TCGCCTCCCGCGGACCAAACTGCAATCACTCCGCCACCCACCAACAACTGGTCT 1855
QY 476 -----LysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaA 492
Db 1856 TCCCTTAAGCAAACTTCACGCTGATCGTCGCGCGGAAAGTCGSCATTGCGGCGCT 1915
QY 492 rgAspAsnAla---AspValProSerGlyLeuMetPheArgSerGlyAlaSerSerV 511
Db 1916 ACGGAGAACCAAGAAATCCCTCTTTGAAACCTTCTACGGCGCGGCGCTGGGTTGCG 1975
QY 511 alArgGlyTyrGluLeuAspSerIleGly----- 520
Db 1976 TCGCGGATACGAAAGCGGACGCTCGTCCGAAAGTGATGACGAATACGGCGAAAAA 2035
QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
Db 2036 TCAGCTACGGCGGCAAC-----AAAAAGCCAAACGCTCCCGCGAGCTGC 2080
QY 540 luTyrGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
Db 2081 TCTTCCGATGCCGCGCGGAAAGACGCGCACCGCTGAGCCTGTTTGGCGAGC 2140
QY 557 etGlyAsp----- 559
Db 2141 CAGGACGCTGTGGACGGCAAAACCTACGACGACACAGACAGTTCGCGACCGGCGCA 2200
QY 560 -----AlaAlaAlaAsnPheLysArgMet-----LysLeuL 570
Db 2201 GGTTCAAAACATTTACGGCGCGGCAATACCCATAATCCACCTTTACCAACGAATTC 2260
QY 570 ysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 2261 GCTATTCGCGCGCGCGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2320
QY 590 leAlaTyr-----GlyHisSerAspLysTyrIleArgTyrPheHisIleSerLeuG 606
Db 2321 ACGCTACCGCTGAAGAAAAACCGGAAGACGAATCCAAACGCTTCCAATTCCAACTCG 2380
QY 606 lyThrArgPhe 609
Db 2381 GCACGACGCTC 2391
RESULT 16
AAH42130
ID AAH42130 standard; DNA; 2394 BP.
XX
AC AAH42130;
XX
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a Neisseria serogroup A protein.
XX
KW Serogroup A protein; outer membrane protein; Neisserial infection;
XX vaccine; ss.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2394
FT /tag= a
FT /product= "Neisseria serogroup A protein"
FT sig_peptide 1..63
FT /tag= b
FT mat_peptide 64..2391
FT /tag= c
XX
PN WO2001:2885-A1.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-1B00166.

```

XX 17-JAN-2000; 2000GB-0001067.
 PR 09-MAR-2000; 2000GB-0005699.
 XX (CHIR-) CHIRON SPA.
 XX Pizza M, Rappuoli R, Giuliani M;
 XX WPI: 2001-451895/48.
 DR P-PSDB; AAB84746.
 XX
 PT Composition for treating or preventing infection to, detecting, or for
 PT raising antibodies against Neisseria bacteria, comprises an N.
 PT meningitidis serogroup B outer membrane preparation and an immunogenic
 PT component -
 XX
 PS Disclosure; Page 70-71; 83pp; English.
 XX
 CC The present sequence encodes a Neisseria serogroup A protein. The
 CC protein is used to produce the compositions of the invention. The
 CC specification describes a composition, comprising a Neisseria
 CC meningitidis serogroup B outer membrane preparation and an immunogenic
 CC component. The immunogenic component is protein disclosed in WO99/57280,
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
 CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
 CC a medicament for treating or preventing infection due to Neisseria
 CC bacteria; a diagnostic reagent for detecting the presence of Neisseria
 CC bacteria or of antibodies raised against Neisseria bacteria; and/or
 CC a reagent which can raise antibodies against Neisseria bacteria. It may
 CC also be used as a vaccine.
 XX

SQ Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;

Alignment Scores:

Pred. No.: 1,39e-13 Length: 2394
 Score: 253.50 Matches: 151
 Percent Similarity: 36.50% Conservatives: 99
 Best Local Similarity: 22.04% Mismatches: 276
 Query Match: 7.99% Indels: 159
 DB: 22 Gaps: 27

US-09-857-669-2 (1-609) x AAH42130 (1-2394)

QY 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
 DB 470 TCGCCGCAACCGCGTCGACATCGACATTCAGAGGCAAAATCCGCCAAATCA 529
 QY 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
 DB 530 CCGACATCGAATTTGAGGCAACCAAGTCTATTCCGACCCCAAACTGATCGGCAGATGT 589
 QY 73 oLeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAgl 93
 DB 590 CGCTGACGAAGCGGCATTTGGACATGGCTGACACAGCAAGCAACCAATTCACAGCAGAGA 649
 QY 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyPhe---- 109
 DB 650 AATTTCCCAACACATCGAAAGAGTAACCGACTTCTACACAGACAACGCGCTACTTCGATT 709
 QY 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaIleThrValHisI 126
 DB 710 TCGGCATCTCGATACCGACATCCAAACCAACGAGGACAAACCAAGCAGACACATCAAAA 769
 QY 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
 DB 770 TCACCGTCCACGAAGCGGACGCTTTCCGTTGGGCAAGTCTCCATC----- 816
 QY 144 spileLeuSerAspGlyAsnLeuAlaGluTyTyArgAsnAlaLeuGlu----- 160
 DB 817 -----GAAGCGACCAACCAAGTTCGCCCAAGCCGCAACTGGAGAAACCTGCTGA 865
 QY 161 -----AsnTrpGlnGlnProValGlySerAspGlnAspSerTrpG 176

DB 866 CCATGAAGCCCGCAAAATGGTACGAA-----C 892
 QY 176 luAsnSerLysThrSerValLeuGlyAlaValAlaThrArgLys-----GlyTyP 192
 DB 893 GCCAGCAGATGACCGCGCTTTGGGTGAGATTCAGAACCGCATGGGTGGCAGGTAGG 952
 QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
 DB 953 CATACAGCGCAATC---AGCGTACAGCGCGTGCACAAAGCCGCAACCAACCGTCGATT 1009
 QY 212 euAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
 DB 1010 TCGTCTCGCACATCGAACCGCGCGGAAATCTACGTCAACAGAAATCCACATCAGCGCA 1069
 QY 232 hrGlnArgTyProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
 DB 1070 ACNACAAACCCCGCAGAGTCTGCGCGCGAATTCGCCCAATTCGAATTCGCGCCCTT 1129
 QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTySerG 272
 DB 1130 ACACACACCTCCAGGCTGCACAGCTCCAAAGAGCGCTCGAGCTTTTGGGCTACTTCGAC- 1188
 QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
 DB 1189 -----ACGCTACAGTTTGTATCCGCTCCGCTTCGCGCACACCGCACAAAGTCGATTGA 1243
 QY 290 ysValSerValThrGluValLysArgHsLysLeuGluThrGlyIleArgLeuAspSerG 310
 DB 1244 ACATGAGCGTCGACGAAGCTTCCACCGCGCTCGCTCGAGCTTCAGCGGGGTGGGTACAG 1303
 QY 310 luTyArgLeuGlyGlyLysIleAlaTyArgTyTyAsnLeuPheAsnLysGlyTyI 330
 DB 1304 ATACCGCGCTGCTATGTCGCGAGCGGTTTCCCAAGACAACCTGTTCGGTACGGGC---- 1359
 QY 330 leGlySerValValTrpAspMetAspLysTyGluThrLeu----- 344
 DB 1360 --AAGTCGGCGCCCTGCGCGCTCCACGAAGCAAAACACGCTCAACGCGCTCGCTGCTGT 1417
 QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyArgGlyAsnT 358
 DB 1418 TTACCGACCCGTACTTTCAGCGCAGACGGGTGTCAGCTGGGTACGATGTTACGGGAAAG 1477
 QY 358 yrTrp-----ThrSerAsnValSerTyAsnArgSerThrThrGlnAsnL 373
 DB 1478 CCTTCGACCCCGCAAGACATCGACACAGCATCAACCAATATAAACCCACCGCAGCGG 1537
 QY 373 euLulysArgAlaPheSerGlyGlyIleTrpTyValArgAspArgAlaGlyIleAspA 393
 DB 1538 CAGGCATCCGCATGAGCGTCTGTACCGAATACGACGCGTGAATTCGGTTCGGTGG 1597
 QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
 DB 1598 CAGAACACCTGACCTCAACACCTACAAACGCGCCCAACCACTATGCGGACTTTATCA 1657
 QY 401 laGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetL 421
 DB 1658 AGAATACGGCAAAACCGCAGCGCAGAC-----GCGACGCTCAAGAGGTGGCTGT 1708
 QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnG 439
 DB 1709 ACAAGGTACCGCTCGGCTGGGGCGCAACAAACCGACGCGGTATGCGCGCGCGCG 1768
 QY 439 lyHisTyLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
 DB 1769 GC---TACCTGACGGGC-----GTCAACGCGCGAAA 1795
 QY 459 euIleArgThrSerAlaArgAlaGlyTyPhePheThrProGluAsnLys----- 475
 DB 1796 TCGCCCTCGCGCGCAACATGCAATACTCCGCCACCCCAACCAACCTGCTTCT 1855
 QY 476 -----LysLeuGlyThrPheIleLeuArgGlnAlaGlyTyThrValAlaA 492

QY 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
Db 1147 CACCTTCCAAAGAGCGGTCGAGCTTTGGCTACTTCGAC-----AATGTCAGTTT 1200
QY 278 AspPheAspArgLeuGlnGly-----AspArgValProValLysValSerValThrGlu 295
Db 1201 GATGCTGTCGCGCTTGCAGCGCCGACAAAGTGAATTTGAACATGAGTGCACCGAA 1260
QY 296 ValLysArgHisLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyCly 315
Db 1261 CGTCCACCGGTTCCCTGGATTTGACGGCGGTTGGGTTCAAGATACCGGTTGGTCATG 1320
QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
Db 1321 TCCGCGAGCGTTTCCCAAGACAACCTGTTCCGTACGGGC-----AAGTCGGCGCCACTG 1374
QY 336 AspMetAspLysTyrGluThrThrLeu----- 344
Db 1375 CGCGCTCCAGGACCAACACGCTTTAAGCGCTCGCTGCTTTACTGACCGTACTTC 1434
QY 345 ---AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp----- 359
Db 1435 ACGGACAGCGGGTCAGCTGGGCTAGCATGTTTACGAAAGACCTTCGACCGCGCAAA 1494
QY 360 ---ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPhe 378
Db 1495 GCATCGCACCATCAACCAATATAAACCCACCGAGCGGCGAGGCGATCGCATGAGC 1554
QY 379 SerGlyIleIleTyrValArgAspArgAlaGlyIleAspAla----- 393
Db 1555 GTGCGCTGTACCGAATACGACCGCGCAATTTCCGTTTGGTGGCAGAACACCTGACCGTC 1614
QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 1615 AACCTACACAAGCGCCCACTATGCGGCTTTATCAAGAAATACGGCAAAACC 1674
QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSer 424
Db 1675 GACGGCAGAC-----GGCAGCTTCAAGGCTGGCTGTACAAAGGTACCGTCGCG 1725
QY 425 TrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGly 444
Db 1726 TGGGGGCGACAAACCGACGCGCTTATGCGCGACCGCGCGC---TACCTGACGGCG 1782
QY 445 LysIleGlyThrThrLeuGlyThrPheLeuSerSerAlaLeuIleArgThrSerAla 464
Db 1783 -----GTGACGCGCGAAATCGCGCTGCTGCGCAGC 1812
QY 465 ArgAlaGlyTyrPhePheThrProGluAsnLys-----LysLeu 477
Db 1813 AACCTGCAATACTACTCGCCACCCACCAACCACTGTTCTTCCCTGACGAAACCC 1872
QY 478 GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAla---Asp 496
Db 1873 TTCAGCTGTATGCTCGCGCGCAAGTCGCGCATTTGCGGGCGGTACGGCAGAACCAAGAA 1932
QY 497 ValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeu 516
Db 1933 ATCCCTCTCTTTGAAACTTCTACGCGCGCGGCTGGTTCGGTGGCGGATACGAAAGC 1992
QY 517 AspSerIleGly-----LeuAlaGlyProAsn 525
Db 1993 GCGAGCTCGGTCGGAAGTCTATGACGAATAGCGGGGAAAAAATCAGCTACGGCGGCAAC 2052
QY 526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro--- 544
Db 2053 -----AAAAGCCACGCTCCGCGGAGCTGCTCTTCCGATGCCCGC 2097
QY 545 -----PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
Db 2098 GCGAAAGACGCGCGCACCTCGGCTGAGCCTGTTTGGCGAGCGAGCGGCTGTGGGAC 2157
QY 559 ----- 559

Db 2158 GCGAAAACTACGACGACACAGCAGGTTCCGCGACCGCGGCGGCTTCAAAACATTTAC 2217
QY 560 AlaAlaAlaAsnPheLysArgMet-----LysLeuLysHisGlySerGlyLeu 575
Db 2218 GCGCGCGCAATACCAATAATCCACCTTTACCAACGAATTCGCTATTTCGCGCGCGCGC 2277
QY 576 GlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAlaTyr----- 592
Db 2278 GCGGTTACTCGCTCGCTTTAGCCCGCATGAAATCAGCTACGCTTACCCGCTGAAG 2337
QY 593 ---GlyHisSerAspLysLysIleArgTyrHisIleSerLeuGlyThrArgPhe 609
Db 2338 AAAAAACCGAAGACGAATCCAAACGCTTCCAATTCCAACTCGGACGACGTTTC 2391
RESULT 18
AAH42128
ID AAH42128 standard; DNA: 2394 BP.
XX
AC AAH42128;
XX
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a *Neisseria* serogroup B protein.
XX
KW Serogroup B protein; outer membrane protein; *Neisseria* infection;
KW vaccine; ss.
XX
OS *Neisseria meningitidis*.
FH Key Location/Qualifiers
CDS 1..2394
FT /*tag= a
FT sig_peptide 1..63
FT /*tag= b
FT mat_peptide 64..2391
FT /*tag= c
XX
PN WO200152885-A1.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-IB00166.
XX
PR 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX
PA (CHIR-) CHIRON SPA.
PI Pizza M, Rappuoli R, Giuliani M;
XX
XX WPI; 2001-451895/48.
DR P-PSDB; AAB84744.
XX
XX Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against *Neisseria* bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component -
XX
PS Disclosure: page 58-59; 83pp; English.
XX
XX The present sequence encodes a *Neisseria* serogroup B protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a *Neisseria*
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO99/57280,
CC WO99/30544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
CC a medicament for treating or preventing infection due to *Neisseria*
CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*
CC bacteria or of antibodies raised against *Neisseria* bacteria; and/or

04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_26 SEQ ID NO:26.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 XX WO200022430-A2.
 XX
 XX 20-APR-2000.
 XX
 XX 08-OCT-1999; 99WO-US23573.
 XX
 XX 09-OCT-1998; 98US-0103794.
 XX 30-APR-1999; 99US-0132068.
 XX
 XX (CHTR) CHIRON CORP.
 XX
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 XX WPI: 2000-318079/27.
 DR
 XX
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 PS Claim 7: Page 532-547; 1760pp: English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SX Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;
 SX
 Alignment Scores:
 Pred. No.: 1.04e-11 Length: 52253
 Score: 253.00 Matches: 134
 Percent Similarity: 36.25% Conservative: 90
 Best Local Similarity: 21.68% Mismatches: 246
 Query Match: 7.97% Indels: 148
 DB: 21 Caps: 25
 US-09-857-669-2 (1-609) x AAA81478 (1-52253)
 QY 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPhe-----SerSer 111
 DB 47964 GAAAAGTAACCGACTTCTACCAAAATACGGCTACTTCGATTTCGGTATCCTCGATACC 48023

QY 112 LysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr-----ProGly 129
 DB 48024 GACATCCAAACCAAGCAAGACAAAACCAAGCAGACCATCAAAATACCTCCACGAAGGC 48083
 QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
 DB 48084 GGAGCTTTCGGTGGGCAAGTCTCCATC-----GAAGGC 48119
 QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu-----Asn 161
 DB 48120 GACACCAACGAAGTCCCAAGCCGAACCTGGAATACTGCTGACCATGAAGCCCGCAAA 48179
 QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
 DB 48180 TGGTAGCA-----CGCCAGCAGATGACCGCC 48206
 QY 182 ValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLysLeuGly 197
 DB 48207 GTTTTGGGTGAGATTTCAGACCGCGCATGGGCTCGGCGAGCTACCATACAGCGAAATC--- 48263
 QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAsp 217
 DB 48264 AGCGTAGACCGCTCCCGCAACGCTGAACCAACCAACCGTCGATTTCTGCTGCACATCGAA 48323
 QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 DB 48324 CGGGGCCGGAATCTACGTCAACGAAATACATACACCGGCAACCAACACCGCGAC 48383
 QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
 DB 48384 GAAGTCGTCGGCGTGAATTACGCCAAATGGAATCCGCACCTTACGACACCTCCCAAGCTG 48443
 QY 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla 277
 DB 48444 CAACGTTCCAAAGACGGCTGCGAGCTTTGGGTACTTTCGAC-----AATGTCCAGTTT 48497
 QY 278 AspPheAspArgLeuGlnGly-----AspArgValProValLysValSerValThrGlu 295
 DB 48498 GATGCTGTCCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 48557
 QY 296 ValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
 DB 48558 CGTTCACCGGCTTCCTGGATTTGAGCGCGGCTGGGTTCAAGTACCGGGTGGTGCATG 48617
 QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrp 335
 DB 48618 TCCGACAGGCGTTTCCCAAGACACGCTGTTCCGTTACGGGC-----AAGTCGCGGCGC 48671
 QY 336 AspMetAspLysTyrGluThrThrLeu-----344
 DB 48672 CGCGCTCCAGGAGGCAAAACACCGCTTAACGGCTCGCTGCTGTTTACTGACCGGTACTTC 48731
 QY 345 ---AlaIaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp-----359
 DB 48732 AGCGACAGCGGGTTCAGCTGGGCTAGCATGTTTACGAAAGCGCTTCGACGCGGCGCAAA 48791
 QY 360 ---ThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPhe 378
 DB 48792 GCATCGACCGCATCAACAATATAAACACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 48851
 QY 379 SerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAla-----393
 DB 48852 GTGCCTGTTTACCGAATAGCAGCGGTGAATTTTCGGTTTGGTGGCAGAACACCTGACCGCTC 48911
 QY 394 -----ArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
 DB 48912 AACACCTACAAACACCG 48971
 QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeu-----ThrAlaSer 424
 DB 48972 GACGGCACACAG-----GGCAGCTTCAAGAGCGTGGCTGTACAAAGGTACCGTCGCGC 49022


```
QY 425 TtpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisThrLeuAspGly 444
DB 49023 TGGGGCGGCACAAACAGCAGCGGCTTATGGCCGAGCGCGGC---TACCTGACGGGC 49079
QY 445 LysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuLeuIleArgThrSerAla 464
DB 49080 -----GTGAAGCGCGAAATCGCCCTGCCTGGCAGC 49109
QY 465 ArgAlaGlyThrPhePheThrProGluAsnLys-----LysLeu 477
DB 49110 AAACGTGCATACTACCTCGCCACCCACCAACCAACCTGTTCTTCCCTTGAGCAAAACC 49169
QY 478 GlyThrPheIleIleArgGlyGlnAlaGlyThrThrValAlaArgAspAsnAla---Asp 496
DB 49170 TTCACGCTGATGCTCGGCGGGAAGTCGGCATTTCCGGCGGCTACGGCAGAACCAAGAA 49229
QY 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyThrGluLeu 516
DB 49230 ATCCCTCTCTTTGAAAACCTTACGCGCGGCGGCTGGTTCGGTGGCGGATACGAAAGC 49289
QY 517 AspSerIleGly-----LeuAlaGlyProAsn 525
DB 49290 GGCACGCTCGTCCGAAAGTCTATCAGCAATACGGCGAAATAATCAGTACGGCGCAAC 49349
QY 526 GlySerValLeuProGluArgAlaLeuValGlySerLeuGluThrGlnLeuPro--- 544
DB 49350 -----AAAAAGCCAAAGTCTCCCGCGAGCTGCTCTTCCGATGCCCGGC 49394
QY 545 -----PheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp----- 559
DB 49395 GCGAAAGACGCGCGACGCTCGCGCTGAGCCTGTTTCCGACGACGACGCTGTGGGAC 49454
QY 559 ----- 559
DB 49455 GGCAAACCTACGACGACACAGCAGTTCCGCGCGCGGAGGTTTCAAAACATTTAC 49514
QY 560 AlaAlaAlaAsnPheLysArgMet-----LysLeuLysHisGlySerGlyLeu 575
DB 49515 GCGCGCGCAATPACCAATAAATCCACCTTTACCAACGAATTCGGCTATTCCGCGCGCGC 49574
QY 576 GlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaThr----- 592
DB 49575 GCGGTACTGCTCTCGCTTTAGCGCCGATGAAATTCAGCTACGCTACCGCTGAG 49634
QY 593 ---GlyHisSerAspLysIleArgTrpHisIleSerLeuGlyThrArgPhe 609
DB 49635 AAAAAACCGGAAGAGAAATCCAACTTCCAACTCCGACGACGAGCTTC 49688
RESULT 20
AAF21544/c
ID AAF21544 standard; DNA: 349980 BP.
AC AAF21544;
XX
XX
DT 13-MAR-2001 (first entry)
DE
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
XX Neisseria meningitidis.
XX
XX WO200066791-A1.
XX
XX 09-NOV-2000.
XX
XX 08-MAR-2000; 2000WO-US05928.
XX
XX 30-APR-1999; 99US-0132068.
XX 08-OCT-1999; 99WO-US23573.
XX 28-FEB-2000; 2000GB-0004695.
```

XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX WPI: 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
PS
XX Claim 7: Appendix A: 692pp: English.

XX The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21598 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21599 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

Alignment Scores:
Pred. No.: 1,48-10 Length: 349980
Score: 253.00 Matches: 134
Percent Similarity: 36.25% Conservative: 90
Best Local Similarity: 21.68% Mismatches: 246
Query Match: 7.97% Indels: 148
DB: 21 Gaps: 25

US-09-857-669-2 (1-609) x AAF21544 (1-349980)

QY 97 AspAsnValLysThrMetLeuArgSerLysGlyThrPhe-----SerSer 111
DB 180140 GAAAAAGTAAACCGACTTCTACCAAAATAACCGCTACTCGATTTCGTATCTCGATACC 180081
QY 112 LysValSerLeuThrGluLysAspGlyAlaThrValHisIleThr-----ProGly 129
DB 180080 GACATCCAAACCAACGACAGACAAACCAACGACAGACCATCAAAATCACCGTCCAGAGCC 180021
QY 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
DB 180020 GGACGTTTCGTTGGGGCAAGTCTCCATC-----GAAGGC 179985
QY 150 AsnLeuAlaGluThrTyArgAsnAlaLeuGlu-----Asn 161
DB 179984 GACACCAACGAAGTCCCAAGCCGAACCTGGAAAAAATCGTGCACATGAAGCCGCAAA 179925
QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
DB 179924 TGTACGAA-----CGCCAGCAGATGACCGCC 179898
QY 182 ValLeuGlyAlaValThrArgLys-----GlyThrProLeuAlaLysLeuGly 197
DB 179897 GTTTTGGGTGAGATTCCAGAACCGCATGGCTCGCAGCGCTACGCATACACCGAATC--- 179841

CC symptomatic disease. They are also useful for the detection of
 CC hybridisation complexes. Antigens and antibodies specific omp proteins
 CC also provide diagnostic, therapeutic and prophylactic compositions for
 CC the treatment or prevention of the infections described above. The
 CC antibodies are useful for inducing a protective immune response in
 CC humans or animals with N. gonorrhoeae, N. meningitidis, or other
 CC Neisseria species. The proteins, antibodies and polynucleotide
 CC sequences of the present invention may also be used in the screening
 CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 2394 BP: 656 A; 752 C; 587 G; 399 T; 0 other;

Alignment Scores:

Pred. No.: 2,13e-13 Length: 2394
 Score: 251.50 Matches: 151
 Percent Similarity: 36.50% Conservative: 99
 Best Local Similarity: 22.04% Mismatches: 276
 Query Match: 7.93% Indels: 159
 Gaps: 27

US-09-857-669-2 (1-609) x AAA15156 (1-2394)

QY 41 SerProAspThrGluSerValLysLeuLys-----ProLysPhe 53
 Db 470 TCGCGCGCAACCGCTCGACATCGACATCGAGGCGCAATCGCGCAAAATCA 529
 QY 54 ProVal-ArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPr 73
 Db 530 CCAGATCGAATTTGAAGGCAACCAAGTCTATTCCGACCGCAAACTGATCGGCGAGATGT 589
 QY 73 oLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAcl 93
 Db 590 CGCTGACCGCAAGCGCGCATTTGGACATGGCTGACACGACGACCAATTCACGAGCAGA 649
 QY 93 uGluAlaProAspAsnValLysThr-----MetLeu-ArgSerLysGlyTyPhe---- 109
 Db 650 AATTTGCCCAAGACATGGAAAGTAAACCGACTTCTACCAAGCAACCGGCTACTTCGATT 709
 QY 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyThrValHisI 126
 Db 710 TCGGTATCTCGATCCGACATCCAAACCAACGAGCAAAACCAAGCAGCACCATAAA 769
 QY 126 leThr-----ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyA 144
 Db 770 TCACCGTCCACGAAGCGGAGCTTTCGTTGGGGCAAAAGTCTCCATC----- 816
 QY 144 spileLeuSerAspGlyAsnLeuAlaGluTyTyArgAsnAlaLeuGlu----- 160
 Db 817 -----CAAGCGCAGACCAACCAAGTCCCAAGCGCAACTGGAAAACTCTCTGA 865
 QY 161 -----AsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpG 176
 Db 866 CCATGAAGCCCGCAATGGTACGAA-----C 892
 QY 176 luAsnSerLysThrSerValLeuGlyAlaValThrArgLys-----GlyTyP 192
 Db 893 GCCAGCAGATGACCGCGCTTTGGTGAGATTTCAGACCGCATGGGCTCGGAGGCTAGC 952
 QY 192 roLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspL 212
 Db 953 CATACAGCGAAATC---AGCGTACAGCGCTGCCCAACCGCGCAAAACCAACCGTCGATT 1009
 QY 212 euAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyT 232
 Db 1010 TCGTCTCGCATCGAACCGGCGCGGAAATCTACGTCACAGAAATCCCATCATCCCGGCA 1069
 QY 232 hrGlnArgTyProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProT 252
 Db 1070 ACACAAACCAACCGCGCAAGTTCGTCGCGCGCGAATTCGCGCAAAATGGAAATCCGCCCTT 1129
 QY 252 yrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTySerG 272
 Db 1130 ACAGACCTCCAAAGCTGCACACCGCTCCAAAGAGCGCGTTCGAGCTTTGGGCTACTTCGAC- 1188

QY 272 lyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----AspArgValProValL 290
 Db 1189 -----AACGTACAGTTTGATGCTCGCGCTTCCGGCAGCACACCGCAAAAGTCGATTTGA 1243
 QY 290 ysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerG 310
 Db 1244 ACATGAGCTGACCGAAGCTTCCACCGGCTCGCTCGACTTGAGCGCGGCTGGGTACAGG 1303
 QY 310 luTyroGlyLeuGlyGlyLysIleAlaTyAspTyTyAsnLeuPheAsnLysGlyTyI 330
 Db 1304 ATACGGCGCTGCTCATGTCCGACGCGCTTCCCAAGACAACCTGTTTCGTTACGGGC----- 1359
 QY 330 leGlySerValValTrpAspMetAspLysTyGluThrThrLeu----- 344
 Db 1360 --AAGTCGGCCCGCTCGCGCTTCACGAAGCAAAACACGCTCAACGGCTCGCTCGCT 1417
 QY 345 -----AlaAlaGlyIleSerGlnProArgAsnTyArgGlyAsnT 358
 Db 1418 TTACCGACCGCTACTTCACGGCAGACGGGTCAGCGCTGGGCTACGATGTTTACGGAAG 1477
 QY 358 yrTrp-----ThrSerAsnValSerTyAsnArgSerThrThrGlnAsnL 373
 Db 1478 CTTTCGACCGCGCAAGCATCGACGACATCAACAATATAAAACCAACCGCAGCGCG 1537
 QY 373 euGluLysArgAlaPheSerGlyGlyIleTyTyValArgAspArgAlaGlyIleAspA 393
 Db 1538 CAGGCATCCGATGAGCGTGCCTGTTACCGAATACGACGCGTGAATTTGGTTCGCTG 1597
 QY 393 la-----ArgLeuGlyAlaGluPheLeuA 401
 Db 1598 CAGACACCTGACCGCTACACCTACAAAGCGCCCAACACTATGCGGACTTTATCA 1657
 QY 401 laGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMeL 421
 Db 1658 AGAAATACGGCAAAACCGCAGCGCACAGAC-----GGCAGCTTCAAAAGCTGCTGT 1708
 QY 421 eu-----ThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnG 439
 Db 1709 ACAAGGTACCGTGGCGGCGCAACAAACCGACAGCGGCTTATGCCCGCAGCGCG 1768
 QY 439 lyHisTyroLysAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaL 459
 Db 1769 GC-----TACCTGACGGC-----GTGACGCGCGAA 1795
 QY 459 euIleArgThrSerAlaArgAlaGlyTyPhePheThrProGluAsnLys----- 475
 Db 1796 TCGCGCTCGCGCAGCAAACTGCAATACTACTCGCGCACCCACCAACCAACCTGCTTCT 1855
 QY 476 -----LysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyThrValAlaA 492
 Db 1856 TCCCTCTTAAGCAAAACCTTCACGCTGATGCTCGCGCGCAAGTCGCAATTCGCGGCGGT 1915
 QY 492 rgAspAsnAla--AspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerV 511
 Db 1916 ACGCGAGAACCAAGAAATCCCTCTTTGAAAACCTTCTACGGCGCGCGCTGGGTTCG 1975
 QY 511 aArgGlyTyroGluLeuAspSerIleGly----- 520
 Db 1976 TCGCGGATACGAAAGCGCGCAGCTCGCTCGCGCAAGTGTATGAGAAATACGCGCAAAAA 2035
 QY 521 --LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuG 540
 Db 2036 TCAGCTACGGCGCAAC-----AAAAAGCAACGCTCTCCGCGCAGCTGC 2080
 QY 540 luTyroGlnLeuPro-----PheThrArgThrLeuSerGlyAlaValPheHisAspM 557
 Db 2081 TCTTCCGATGCTGGCGGAAAGACGCGCGCTCGCTCGCGCTGAGCTGTTTGGCGAGC 2140
 QY 557 etGlyAsp----- 559
 Db 2141 CAGGCGCGTGTGGGACGCGCAAAACCTACGACGACACAGCAGTTTCGCGACCGCGCGCA 2200

```

Qy 560 -----AlaAlaLaAsnPheLysAlqMet-----LysLeuL 570
Db 2201 GGGTTCAAAACATTTACGGCGCGCAATACCCATAATCCACCTTTACCAACGAATTC 2260
Qy 570 yshisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspI 590
Db 2261 GCTATTCCCGCGCGCGGGTTACCTGGCTCTCGCTTTAGCGCGCATGAAATTCAGGT 2320
Qy 590 leAlaTyr-----GlyHisSerAspLysLysIleAqGTrpHisIleSerLeuG 606
Db 2321 ACGCCATCCCGTGAAGAAAAACCGGAGACGAATCCAAACGCTTCCCAATTCCAACTCG 2380
Qy 606 lyThrArgPhe 609
Db 2381 GCACGACGTTTC 2391

RESULT 22
AAI97964
ID AAI97964 standard; DNA; 6617 BP.
XX
AC AAI97964;
XX
DE 20-NOV-2001 (first entry)
XX
DE Lawsonia Intracellularis coding sequence SEQ ID NO: 1.
XX
KW HtrA; PonA; HypC; YefW; ABC1; Omp100; Lawsonia Intracellularis infection;
XX vaccine; ds.
XX
OS Lawsonia Intracellularis.
XX
PN JP2001169787-A.
XX
PD 26-JUN-2001.
XX
PF 20-OCT-2000; 2000JP-0320735.
XX
PR 22-OCT-1999; 99US-0160922.
XX
PA (PF12 ) PFIZER PROD INC.
XX
PS WPI; 2001-592540/67.
XX
PT Lawsonia Intracellularis polynucleotide and encoded protein, used to
prevent Lawsonia Intracellularis infection -
XX
XX Claim 3; Page 37-39; 67pp; Japanese.
XX
XX The present invention provides isolated polynucleotides encoding HtrA,
CC PonA, HypC, LysS, YefW, ABC1 or Omp100 protein of Lawsonia
CC Intracellularis. The sequences can be used in vaccines for the prevention
CC of Lawsonia Intracellularis infection. The present sequence is a coding
CC sequence of the invention.
XX
XX SQ Sequence 6617 BP; 2216 A; 1153 C; 1142 G; 2106 T; 0 other;

Alignment scores:
Pred. No.: 2,24e-12 Length: 6617
Score: 247.00 Matches: 140
Percent Similarity: 38.21% Conservative: 95
Best Local Similarity: 22.76% Mismatches: 276
Query Match: 7.78% Indels: 104
DB: 22 Gaps: 25

US-09-857-669-2 (1-609) x AAI97964 (1-6617)
Qy 64 IleLysAspMetValGluGluHisLeuProLeuIle-----ThrGlnGlnGlnGlu 80
Db 4643 ATTAAGATGTCGGAATTTGAGGACTTGAACATAAAGCTAAACGCTTTAAAAAAGAG 4702
Qy 81 GluValLeuAspLysGluGln-----ThrGlyPheLeuAlaGlu 93
Db 4703 TTAGCATTAACAGACGTAATTTTATCATGGTGTATTACATGGAACAGGTGATTACGTGAA 4762

```

```

Qy 94 GluAlaProAspAsn-----ValLysThrMetLeuArgSerLysGlyTyrPhe 109
Db 4763 GAATATCTTGAACGGTACTCTATAGCAATCTCTGCCCTATGCCATGAATCATGGCTGTGTA 4822
Qy 110 -----SerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrVal 124
Db 4823 GATATTTCAAGTTGCTTCCCTCACTGAAGTCAATCAATGAAAAAGGAATTTGTTATTACATTT 4882
Qy 125 HisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaLleLeuGlyAsp 144
Db 4883 AGATATAAAGAGAGTAAGCGCTATAAATA-----GGAAAA 4918
Qy 145 IleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeu----- 159
Db 4919 ATAGACTTTAAAGGAGATCTTATTGAGACAAATGAACAACCTCTTAAAGTAACAAAAATT 4978
Qy 160 -----GluAsnTrpGlnProValGlySerAspPheAspGlnAspSerTrpGlu 176
Db 4979 GATGATCATAAAACTATGAGCAGTATTTTCTCTTCTGTATGCAAGATGATGTAAAA 5038
Qy 177 AsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeu 196
Db 5039 -----GCATTAACAGATTTTATTTCAGATTATGCTTATGCTTAAAGTA 5086
Qy 197 GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 216
Db 5087 ---GATCTTGAACAACCAACCAAAATGAAGAATGCAACAATGATGTACTTTCCCTATT 5143
Qy 217 AspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPro 236
Db 5144 GATAAAAAACAAAGTCTTCTTCGTAGATAATTTGTTGAAGGAATACTCTCTACTAGA 5203
Qy 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 256
Db 5204 GATAANTGTTATCTCCGCGAATTACGCCCTCTGCTGATGAGATCTTTTAAATGCTCAACAT 5263
Qy 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
Db 5264 CTCGACGCTCTAATGAATGCTTAAACGCCCTTGCTATTATTTAAC-----CAA 5311
Qy 277 AlaAspPheAspArgLeu-----GlnGlyAspArgValProValLysValSerVal 293
Db 5312 GTAGATACAGATACACTGCTACAGGGAAGATGATGAAGTTGATCTACTTGTAAAGTT 5371
Qy 294 ThrGluValLysArgHisLysLeuGluThrGlyIle-----ArgLeuAspSerGlyTyr 311
Db 5372 CAAGAAGCTCGAACAGGTGCAATCACAGGTGCTGTGTTACTCACACATTTCTAAATTT 5431
Qy 312 GlyLeuGlyGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLysGlyTyrIleGly 331
Db 5432 GGTGTTTCAGGAAGTATCTCA-----CAAGAACAATTATGGGGAAGGTATATTTTA 5485
Qy 332 SerValValTrpAspMetAspLysTyrGluThrThrLeuAlaLaglyIleSerGlnPro 351
Db 5486 AGTATTGAAGGTTTATTTCTAGTAAGTCATCTCTCTCTGATCTTTTTCACCAATCCT 5545
Qy 352 ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsn---ArgSerThrThr 370
Db 5546 CGTGTTTATGATACAGACTTTGGCTTTAGTAATAACATTATATAGCTACAGATGAATGG 5605
Qy 371 GlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspAlaGly 390
Db 5606 GATGACTTCGGTAAAAAAACTTATGGA-----GATACCATACGT 5644
Qy 391 IleAspAlaArgLeuGlyAla---GluPheLeuAlaGluGlyArgLysIle----- 406
Db 5645 CTATTTTACCCCTATAGGAGATATTTCATCTATCTTTGTTGGCTATCGAATTGATCAATAT 5704
Qy 407 -----ProGlySerAspIleAspLeu---GlyAsnSer 416
Db 5705 CGTCTATATGATATTCATCTACAGCACGCTCTTATCTTACTTACTATCAAGGAAAAAT 5764

```



```
Db 1422 -----AACAAATCTCTGGTACACTGGCAC--- 1399
Qy 165 ProValIcylSerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 1399 -----
Qy 185 AlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsn 204
Db 1399 -----
Qy 205 ProAspThrAlaThrValAspLeuAsnVal---ValValAspSerGlyArgProIleAla 223
Db 1398 ---GATACAGCCCGGTCGACGGAAATCTTCATCTAATTGATACACCA----- 1354
Qy 224 PheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 1353 -----CCAGAGGGCCTTATATATACGACC 1330
Qy 244 AlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAla 263
Db 1329 CTCACGCTGCACCA---ACGCCAACTAAC-----AAATCA 1297
Qy 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 1296 ACGCCCAACAACCTTCACCTGGGTGCCGA----- 1267
Qy 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
Db 1266 GGAGAC-----GTAACC-----CGAAACCAACAGAACCGGCCT 1234
Qy 304 GlyLeuArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTrpAsn 323
Db 1233 GGACAGGAGCCAGAAAT----- 1216
Qy 324 LeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThr 343
Db 1215 -----ATGAGCGCCCATGACACTATG 1195
Qy 344 LeuAlaAlaGlyIleSerGlnPro-----ArgAsnTyrArgGlyAsnTyrTrp 359
Db 1194 CTGCAAGGGATGTGACGGCGTCCATACATCTTGGCGCTTCGAGAGAAC----- 1141
Qy 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu-----Glu 374
Db 1140 -----CAGAACTCTGTAGACTCTGACTACT 1114
Qy 375 LysArgAlaPheSerGlyGlyIle-TrpTyrValArgAspArgAlaGlyIleAspAla 394
Db 1113 CAGAAAGGTTTCCCTGGCTGTGTAAGCGTTTCAGTGGCGGCGGTACCCCGTCCGA 1054
Qy 394 gLeu-----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAs 412
Db 1053 TTGTGTAATCGGAGGAGTGTGTGGCA-----CGGCCTGGCTTATCTCGACGCTGCT 1000
Qy 412 pLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAs 432
Db 999 GATC-----ACCTTCTCTGCATTTCAAACCGCGCGCTGGC-AATCA 956
Qy 432 nValLeuHisProGluAsnGly----- 439
Db 955 TGTGCTCATTCGCAAGCGCATTTGACGTTACTTCCCTTCCAGAACCCCGACATCACAC 896
Qy 440 -----HisTyrLeuAspG1 444
Db 895 GCAAGCGGCTGGCAGCGTGCATTAACCTCGCGTGGAGTCTCGACCACTTTACTCAGG 836
Qy 444 yLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAl 464
Db 835 TGAATATACCAATACCAGCAGTCTGTTTATCTCTGGGGTATGATTAGCGCACGGCTTC 776
Qy 464 aArgAlaGlyTyrPhe----- 469
Db 775 TCGTGTGGCTGATGCCNAACCTGGGGGCGACTCGCAACGCTACTCTATCTACTCTCCAA 716
```

```
Qy 470 -----PheThrProGluAsnLysLysLeuG1 478
Db 715 CACGGGCTGGGTTCAGATGTCGATTTCTCGTTTTCAGGCGCAGAACCTCTGGATCCG 656
Qy 478 yThr-----PheIleIleArgGlyGlnAlaGlyTyrThrValAlaAr 492
Db 655 CACACTGTACGATCGCCATCGTTTGTTCACAGCGGCACGCTGGCTGGATTGAACCCG 596
Qy 492 gAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValAr 512
Db 595 TGATTTTCGACAAAGTACCCCGGATCTCGCTTCTTCCCGGGGGCGACCGCATTTTCG 536
Qy 512 gGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluAr 532
Db 535 TGCGTCAAAATCAAAATCTATCGCTCCCGAAATACGCCAAGTGCACCTGAAGGGGCTC 476
Qy 532 gAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAl 552
Db 475 GAAGTTTGATAACCGGATCGCTCGAATACCACTACACGTTGACCGGAAATGGTGGGCGC 416
Qy 552 aValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeuLysHisG1 572
Db 415 GGTGTTTGTGATAGTGGCGAAGCGGTAAAGCATATTCGCGCAGCGACTTTAAACCCG 356
Qy 572 ySerGlyLeuGlyValArgTrp-PheSerProLeuAlaProPheSerPheAspIleAlaT 592
Db 355 TACCGGGTGGCGTGGCTGGGAATGCGCGCTCGGCCCAATCAAACTCGATTTTCCG 296
Qy 592 yRGlyHisSerAspLysLys-----IleArgTrpHisIleSerLeuGly 606
Db 295 TACCGGTGCGGATAAAGAACACGCGTTACAGTTTACATCGCTCGTGGG 243

RESULT 24
ABQ72980
ID ABQ72980 standard; DNA; 2388 BP.
XX
AC ABQ72980;
XX
DT 20-SEP-2002 (first entry)
XX
DE Methylococcus capsulatus surface protein-antigen D15 DNA SEQ ID NO:5.
XX
KW Methylococcus capsulatus; methanotropic bacterium; MopC; MopD; MopE;
KW MopF; D15; surface protein antigen; helper protein; vaccine;
KW epitope mapping; gene; ds.
XX
OS Methylococcus capsulatus.
XX
PN W0200255549-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002NO-NO00018.
XX
PR 12-JAN-2001; 2001NO-0000238.
XX
PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX
PI (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557813/59.
XX
PT New nucleotide molecule encoding a surface exposed protein and is
PT incorporated and expressed in Methylococcus capsulatus, useful in
PT manufacturing vaccines that can be administered orally in animals, fish
PT or humans
XX
PS Claim 1; Page 34-35; 39pp; English.
XX
```

CC The present invention describes a nucleotide molecule (1) comprising a
 CC sequence that codes for any of the 14 *Methylococcus capsulatus*
 CC nucleotide sequences given in ABQ72976 to ABQ72989. Also described is
 CC a method for producing a desired protein in a bacterial host cell,
 CC comprising: (a) transforming a bacterial host cell with a recombinant
 CC vector comprising a first nucleotide sequence selected from the
 CC group of 14 *Methylococcus capsulatus*, and comprising a nucleotide
 CC encoding the desired protein linked in frame to the first nucleotide
 CC sequence; and (b) culturing the transformed host cell in a medium under
 CC conditions that allow the expression of the protein. (1) can be used
 CC in vaccine production. The nucleic acids or the proteins from the
 CC present invention can be used in manufacturing vaccines that can be
 CC administered in animals, fish or humans. The method is useful in
 CC producing fusion proteins carrying foreign peptide sequences that can
 CC also be useful in epitope mapping, screening of antibody libraries and
 CC as immunogens. The present sequence represents the specifically claimed
 CC *Methylococcus capsulatus* surface protein-antigen D15 nucleotide sequence,
 CC from the present invention.

XX
 SQ Sequence 2388 BP; 535 A; 653 C; 686 G; 514 T; 0 other;

Alignment Scores:

Pred. No.:	8.le-12	Length:	2388
Score:	234.50	Matches:	151
Percent Similarity:	32.98%	Conservative:	99
Best Local Similarity:	19.92%	Mismatches:	259
Query Match:	7.39%	Indels:	249
DB:	24	Gaps:	29

US-09-857-669-2 (1-609) x ABQ72980 (1-2388)

QY	20	TyrAlaProAlaAlaAsp-----	LeuSerGluAsnLysAlaAlaClyPhe-----	34
DB	229	TACCTGCCTGTCCGGGAGGCGACCCCTCGATGAAAAACGGTCGCCGAGGTATCCGG		288
QY	35	AlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhePro		54
DB	289	GGCGTGTTCAG-----	ACGGGTTTTTTCAG-----	318
QY	55	ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu		74
DB	319	GTCCGCTGTGGACGAGGACGCGGTAGTTGATCATCTAGCTCGAAGACGGT---CCTTCG		375
QY	75	IleThrGlnGlnGln-----	GluGluValLeuAspLys	85
DB	376	ATTTCCAGCGTAAAGATCGACGCCAACCATGACATCGGTAGCGAGGATCTGCTGAAGCG		435
QY	86	GluGlnThrGlyPheLeuAlaGlu-----	GluAlaProAspAsnVal	99
DB	436	CTCAAGGAATCGGCTGGCGGAGGAGGATTCGACCGGCAGATTCGACAAAGTGC		495
QY	100	LysThrMetLeuArgSerLysGlyTyThrPheSerSerLysValSerLeuThrGluLysAsp		119
DB	496	GAACAGGAACCTCGCGCCAGTATTACAGCGCGGCAATAACAGCCTT-----		543
QY	120	GlyAlaTyThrValHisIleThrProGlyProArgThrLysIleAla---AsnValGly		138
DB	544	---AAATCGATTCCAGGTGACGGAACCTGCCCGGACCGGTGGCGGTGAATCAAT		600
QY	139	ValAla-----	IleLeuGlyAsp-----	144
DB	601	ATCGCGGAGGGCGGTGCTGCATAAAAACAGATCAATATCATCGGAAATAATGCATTT		660
QY	145	-----	IleLeuSerAspClyAsnLeuAlaGluTyThr	155
DB	661	AGTGATGATGATCTGACCCAGGATTTTGACTGAGTACATCCCACTCTCTTCTTCAT		720
QY	156	ArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrp		175
DB	721	ACCAAGGATGATCATGATTCCAGCAAAATTTGCGCGGATCTGGAGCGG-----		771
QY	176	GluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyThrProLeuAlaLys		195
DB	1753	-----	GACCATGGCAGGAGTATTCGACCTATTCGCTCCTCCTCCTGGGTGG	1797

DB	772	-----	CTGCGCTCCTACTCTCGATCGCGGTACGTCAATTTGAA	813
QY	196	LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal		215
DB	814	ATCAGTCCACCCAGGTATCGATCAGCCCAAGAAATATATATACCATCAAC		873
QY	216	ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGly---ThrGlnArg		234
DB	874	GTTAAGAGGCGCGAGGTCTTCAAGGTGGACAGGTGAGCTGACCGCAAGCAACATCTGTG		933
QY	235	TyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeu		254
DB	934	CCACCCGAACAACATCGTCCGG---CTCGTGGTATCGCGCGGAGGATATTTTCCAGG		990
QY	255	AspLeuLeuLeuAspPheGlnGlnAla-----		263
DB	991	AAACTGGCGACAGAAACCCAGAGGGCATTTCCGATCGTTTGGCGGAGGAGGCTATATC		1050
QY	263	-----	-----	263
DB	1051	TTCCGCAACGTCAATATATGTCGCGGACATCAATCAGGAGAAGAACCCGTAACATCACC		1110
QY	263	-----	-----	263
DB	1111	TTTTTCTGATCCGGGCAACAGGTCTAGCTGAGCGGAATCAATTTCCAGGGCAATACC		1170
QY	263	-----	-----	263
DB	1171	AAGACGCTGACGAAGTCTCGCAGGGAATCGGCAGATGAGGCGGCTGGCGCTCG		1230
QY	264	-----	LeuGluGlnAsnGlyHisTyrSerGlyAla	273
DB	1231	ACCGGGAAGATCGAGCTTCCAGACAGGCTCGAGCTCGGCTATTTCCAGGATGTC		1290
QY	274	SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal		293
DB	1291	AACTGGAGAGCCCGCGGTGCGGGAAACAGCATCAGTCGATCACTACAGTGTG		1350
QY	294	ThrGluValLysArgHisLysLeuGluThrGlyLeuArgLeuAspSerGluTyrGlyLeu		313
DB	1351	ACCGAAGTCTTCCGCAACCTGACCGCGGTGTAGGTATTTCCAGTTCACAGGCAATC		1410
QY	314	GlyGlyLysIleAlaTyThrAspTyThrAsnLeuPheAsnLysGlyTyThrIleGlySerVal		333
DB	1411	ATATTCAATCGCGGCTGACCCAGGACAACATTTTCGGCAGCGCAAGCGTGTGATTC		1470
QY	334	ValTrpAspMetAspLysTyThrLeuAlaAlaGlyIleSerGlnPro-----		351
DB	1471	AAATTCACACAGCCAGATCAACACCATCTATGCTTGGGTATTTCAATCCTTACGCC		1530
QY	351	-----	-----	351
DB	1531	ACCTGGATGGATCAGCAGTGTTCACATCAGCTATCGCACACCAACACCGGCTAT		1590
QY	352	ArgAsnTyThrArgGlyAsnTyThr-----		360
DB	1591	TCGAATATGTTCCCACTACATTTACCAATGTTTCCAGTGGCGGCAACTGGGGTTG		1650
QY	361	-----	SerAsnValSerTyrAsn-----	Arg
DB	1651	CCCATCGGAGATTCGACAGTATTCGGACCAACAGCTGACAGTACAGTAAACCAAGCTCAAG		1710
QY	368	SerThrThrGlnAsnLeuLys---ArgAlaPheSerGlyGlyIleThrTyThrValArg		386
DB	1711	ACGACCTCCAGTCTTCGATCAATTCAGACCGTTCATTCGC-----		1752
QY	387	AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle		406
DB	1752	-----	-----	1752
QY	407	ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla---SerTrp		425
DB	1753	-----	GACCATGGCAGGAGTATTCGACCTATTCGCTCCTCCTCCTGGGTGG	1797

```
QY 426 LysArgGlnLeuLeuAsnValLeuHisProGluAsnGly-----HisTyrLeuAsp 443
DB 1798 ACCCATGACACCCCTGAATCGGGCGGATATTCGCCAGCAGCGGTGGCGCGCAGCGATTGACG 1857
QY 444 GlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
DB 1858 GCAGCTGTTTGGCCCTGCGCTTCAGCACCCCTCAGTATTACAAGGCCAACGTGGCCTGGAG 1917
QY 464 AlaArgAlaGlyTyrPhePheThrProGluAsnLysLeuGlyThrPheIleLeuArg 483
DB 1918 CAG-----TATTTT-----CCGCTGACCCAGGATCTG---ACGCTGGTGGTTGAAT 1959
QY 484 GlyGlnAlaGlyTyr-----ThrValAlaArgAspAsnAlaAspValProSerGly 500
DB 1960 GCGGATTTTCGGTTACGGCGCGCGGTATGGCAGTGGCGGCAACAGCGTCTTCCGTTCG 2019
QY 501 LeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
DB 2020 GAACACTTTTACGGCGGGGTCCGATTCCTGCTGGATACGACCCCAATTCACCTGGG 2079
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
DB 2080 CCGAGGACAGCCGTGGCTACGCTTCGGCGGCAACAGCAAACTGACCGGCTCGGTGGAG 2139
QY 541 -----TyrGlnLeuProPheThr-----ArgThrLeuSerGlyAlaValPhe 554
DB 2140 TTGCTGTTTCGGTCCCGCTTCGGCGGAGAACTGACAGTGTCCGCTTGGGTACCTTC 2199
QY 555 HisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLys---LeuLysHisGlySer 573
DB 2200 GTCGACGGCGGAATGCTGCTGCAATTCGCCCGACGTTTCGGACCTGCGCTTCGAC 2259
QY 574 GlyLeuGlyValArgTyrPhePheSerProLeuAlaProPheSerPheAspIleAla 591
DB 2260 GCATTTTCGGCCAGTGGCTATCACCTTCGGGGCCCTGATGTTTCGATCGCC 2313

RESULT 25
ID ABQ90024 standard; DNA; 2391 BP.
XX AC ABQ90024;
XX DT 01-OCT-2002 (first entry)
XX DE M. capsulatus gene #9 for DNA array.
XX KW Micro array; gene; ds; differential expression; gene expression.
XX OS Methylococcus capsulatus.
XX PN WO200255655-A2.
XX PD 18-JUL-2002.
XX PF 14-JAN-2002; 2002WO-N000019.
XX PR 12-JAN-2001; 2001NO-0000235.
XX PR 12-JAN-2001; 2001NO-0000239.
XX PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX PA (TIGR-) TIGR.
XX PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eidsen JA, Fraser CM, Durkin AS;
XX PI Salzberg SL;
XX DR WPI; 2002-557818/59.
XX PT Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes
```

```
XX PS Claim 19; Page 50; 678pp; English.
XX CC The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC invention.
XX SQ Sequence 2391 BP; 536 A; 653 C; 687 G; 515 T; 0 other;

Alignment Scores:
Pred. No.: 8,11e-12 Length: 2391
Score: 234.50 Matches: 151
Percent Similarity: 32.98% Conservative: 99
Best Local Similarity: 19.92% Mismatches: 259
Query Match: 7.39% Indels: 249
DB: 24 Gaps: 29

US-09-857-669-2 (1-609) x ABQ90024 (1-2391)
QY 20 TyrAlaProAlaAlaAsp-----LeuSerGluAsnLysAlaAlaGlyPhe----- 34
DB 229 TACCTGCTGCTCGCGAGGGGACACCCCTCGATGAAACGGTCGCCCGAGGTGATCCGG 288
QY 35 AlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhePro 54
DB 289 GCGCTGTTTCAAG-----ACGGGTTTTTCAAG-----GAC 318
QY 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
DB 319 GTCCCTCTGGACGAGGAGCGGTAAGTTGATCATCTACGTCGACGAGCGT---CCTCG 375
QY 75 IleThrGlnGlnGln-----GluGluValLeuAspLys 85
DB 376 ATTTCCAGCGTAAGATCGACGGCAACCATGACATCGTAGCGAGGATCTGCTGAAGCG 435
QY 86 GluGlnThrGlyPheLeuAlaGlu-----GluAlaProAspAsnVal 99
DB 436 CTCAGGGAATCGGCGTGGCGGAGGAGGATTCGACCGCGGAGATTCTCGACAGGTC 495
QY 100 LysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysasp 119
DB 496 GAACAGGAACCTCGCGCGCATATTACAGCGCGGCGCAATACAGCCTT----- 543
QY 120 GlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAla---AsnValGly 138
DB 544 ---AAATCGATTCCAGGTGACGGAACCTGCCCGAACCGTGTGCGGTGAATATCAAT 600
QY 139 ValAla-----IleLeuGlyAsp----- 144
DB 601 ATCGCGGAGGGGCGGTGCTGCGCATAAACAGATCAATATCATCGGAATATGATTCATT 660
QY 145 -----IleLeuSerAspGlyAsnLeuAlaGluTyrTyr 155
DB 661 AGTGATGATGATCTGACCCAGGATTTTGGCTGAGTACATCCCAATCTCCTCTCTCTAT 720
QY 156 ArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrp 175
DB 721 ACCAGGATGATCAGTATTCACAGCAAAATGTGCGCGCATCTGGAGCGG----- 771
QY 176 GluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLys 195
DB 772 -----CTGCGCTCTACTATCTCGATCGCGGTTACGTCATTCGAA 813
QY 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
DB 814 ATCGAGTCCACCCAGGTGATGATCCGCCCAACAGAAATATATATACCATCCACAC 873
QY 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGly---ThrGlnArg 234
DB 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGly---ThrGlnArg 234
```


Db 874 GTTAAGGAGGCGAGGTCTTCAAGTGGACAGGTGACCGCAAGACCATCGTG 933
QY 235 TyrProGluGlnIleValSerGlyLeuAlaAArgPheGlnProGlyThrProTyrAspLeu 254
Db 934 CCACCCGAACAACTCTGCCCCG---CTGTCGTGATCGGCGCGAGGATATTTTTCACAG 990
QY 255 AspLeuLeuLeuAspPheGlnGlnAla----- 263
Db 991 AAACCTGGCGACAGAAACCCAGAAAGGCATTTCCGATCGTTTGGCGGAGGAGGCTATATC 1050
QY 263 ----- 263
Db 1051 TTTCGCCAACGTCAATATGTTGCCGACATCAATCAGGAGAAGAACCCGTAACATCAC 1110
QY 263 ----- 263
Db 1111 TTTTTCGTGATCCGGGCAACAGGTCTACGTGAGCGGAATCAATTTCCAGGCGCATACC 1170
QY 263 ----- 263
Db 1171 AAGACCGGTGACGAAGTCTCGCGAGGAAATCGGCAGATGAGAGCGGCCCTGGCGTCG 1230
QY 264 -----LeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 1231 ACCGCGAAGATCGAGCGTTCCAGACAGCGCTCGAGGCTCGGCTATTTCCAGGATGTC 1290
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerVal 293
Db 1291 AACGTGGAGAGCGCGCGGTGCCGGGAACGACGATCATGATCGAGCTCAACTACATGCTG 1350
QY 294 ThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeu 313
Db 1351 ACCGGAAGTCTTCGGCAACCTGACGGCGGTGTAGGTATTTCCAGTTCACGAGGCATC 1410
QY 314 GlyGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLysGlyTyrIleGlySerVal 333
Db 1411 ATATTCAATGCGGGGTGACCCAGGACCAATTTTCGGCAGCGGCAACGCTCTCAGTTTC 1470
QY 334 ValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro----- 351
Db 1471 AATTTCAACACAGCGAGATCAACACCATCTATGCTTGGGTATTTTCAATCCTTACGCC 1530
QY 351 ----- 351
Db 1531 ACCCTGGATGGATCAGCAGTGGTTTCGACATCAGCTATCGCGACACCAACACCGGCTAT 1590
QY 352 ArgAsnTyrArgGlyAsnTyrThr----- 360
Db 1591 TCGAACTATGTTCGCAACTACATTACCAATGTGTTCAGTGGCGGCAACTGGGGTTG 1650
QY 361 -----SerAsnValSerTyrAsn-----Arg 367
Db 1651 CCCATCGGAGAGTTCGACAGATTTCGGACCACTGGACTACAGTAACCAACCAAGTCAAG 1710
QY 368 SerThrThrGlnAsnLeuGluLys---ArgAlaPheSerGlyIleTyrIleTyrValArg 386
Db 1711 ACGACCTCCAGTCTTCGGATCAATCAACAGCGCTTCAATGCC----- 1752
QY 387 AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 1752 ----- 1752
QY 407 ProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla---SerTrp 425
Db 1753 -----GACCATGGCAGCGAGTATTCGACCTATTCGTCCTCACTGGGTGG 1797
QY 426 LysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGly-----HisTyrLeuAsp 443
Db 1798 ACCCATGACACCTGTAATCGGGCGATATTCGCGACGAGCGGTGGCGGACGAGTATGACG 1857
QY 444 GlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 1858 GGACTGTTTGGCCTGCCGTTTCAGACACCTGCAGTATTACAAAGGCCAACCTGCGCCTGGAG 1917

QY 464 AlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg 483
Db 1918 CAG-----TATTTT-----CCGCTGACCCAGGATCTG---ACGCTGTGGTTGAAT 1959
QY 484 GlyGlnAlaGlyTyr-----ThrValAlaArgAspAlaAspValProSerGly 500
Db 1960 GCGGATTTTCGTTACGCGCGGTATGCGCAGTGGCGCAACAGCGCTCCTTCGTTCTGG 2019
QY 501 LeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGly 520
Db 2020 GAACACTTTTACGCGCGGTCTCCGAATTCGCTGCGTACGACGCCCAATTCATCTGGGG 2079
QY 521 LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGlu 540
Db 2080 CCCAGGACAGCGCTGCTACGCTTCGCGCGCAACAGCAACTGACCGCTCGTGGAG 2139
QY 541 -----TyrGlnLeuProPheThr-----ArgThrLeuSerGlyAlaValPhe 554
Db 2140 TTGCTGTTTTCGCGTCCCGTTTTCGCGCGAGAAACTCAAGAGTGTCCGCTTGGGTACCTTC 2199
QY 555 HisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLys---LeuLysHisGlySer 573
Db 2200 GTCGAGCGGGAATGTGTCTCAATTCGCCCGCAGCTTTCGGACCTGCGCTTCGACC 2259
QY 574 GlyLeuGlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAla 591
Db 2260 GGCATTTTCGCCAAGTGGCTATCACCTTCGGGGCCCTGATGTTCTCGATCGCC 2313

RESULT 26

AAS81897

ID AAS81897 standard; cDNA; 1425 BP.

XX AAS81897;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #17701.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG17710.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 1: SEQ ID No 17701; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques

Db 1298 GCTGTTATGTGGAGTCCCAATTAAGTACGAGCAGATGCCCCAGTGGTGGCA---GG 1242

Qy 451 yThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheTh 471

Db 1241 ACCATGACAGCCGCCCGGCTATTCTGCGAGCGACGCCAGCAGAGCTCTGGA 1182

Qy 471 rProGluAsnLysLeuGlyThr-PheIleIleArgGlyGlnAlaGlyTyrThrValA 491

Db 1181 TCCGACACTGTACGATGCCATCGTTTGTGTACAGCGGCGACGCTGGCTGGATTGAAA 1122

Qy 491 laArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerS 511

Db 1121 CCGGTGATTTCGAAAGTACCGCGGATCTCGCTTCTTCGCGGGGCGACCGCAGTA 1062

Qy 511 alArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProG 531

Db 1061 TTCGTGGCTACAAATCAATCTATCGTCCGAATACGCCAAGCGTGACCTGAAGGGG 1002

Qy 531 luArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerG 551

Db 1001 CCTCGAAGTTGATAACCGGATCGCTGGAATACCAAGCTACACGTCGCGGAAATGCTGGG 942

Qy 551 lYAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPhelYsArgMetIlysiLeuLysH 571

Db 941 GCGGGTGTTCGTGATAGTGGGAGCGGTAAAGCGATATTCGCGCGAGCCACTTTAAAA 882

Qy 571 lSgLySerGlyLeuGlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleA 591

Db 881 CCGGTACCGGGGTGCGCGTGGCGTGGGAATCGCGGTGCGGCGCAATCAAACTCGATTTTG 822

Qy 591 laTyrGlyHisSerAspLysLys-----IleArgTrpHisIleSerLeu 605

Db 821 CCGTACCGGTGCGGATAGACGACGCGGTACAGCGGTACAGTTTACATCGGACTG 769

RESULT 28

AAAS0537

ID AAAS0537 standard; DNA; 2670 BP.

XX AC AAAS0537;

XX DT 05-DEC-2000 (first entry)

XX DE Moraxella catarrhalis BASB081 mature protein coding region.

XX KW BASB081 gene; infection; otitis media; pneumonia; sinusitis;

XX KW inflammation; therapy; antibacterial; antiinflammatory; vaccine;

XX KW diagnosis; ds.

XX OS Moraxella catarrhalis.

XX FH Key Location/Qualifiers

XX FT conflict replace(1003,G)

XX FT /*tag= a

XX FT /note= *corresponds to base 1093 of AAAS0536*

XX PN W0200052042-A1.

XX PD 08-SEP-2000.

XX PF 23-FEB-2000; 2000MO-EP01468.

XX PR 26-FEB-1999; 99GB-0004559.

XX PA (SMIK) SMITHKLINE BECHAM BIOLOGICALS.

XX PI Ruelle J;

XX DR WPI; 2000-587296/55.

XX DR P-PSDB; AAY95988.

XX PT New BASB081 polypeptides from Moraxella catarrhalis and polynucleotides

XX PT encoding the polypeptides used for treating infections, or as a vaccine

XX PT for preventing infections, especially those caused by M. catarrhalis -

XX Clalm 8; Page 60-61; 97pp; English.

XX The present sequence is that of DNA encoding BASB081 mature protein

CC (see AAY95988) of Moraxella catarrhalis Mc3931 (ATCC 43617). It was

CC obtained by sequencing PCR-amplified DNA from a BASB081 expression

CC plasmid. The sequencing shows 99.9% identity to a mature BASB081

CC coding sequence (see AAAS0537) obtained from M. catarrhalis ATCC 43617.

CC The invention provides BASB081 polypeptides, polynucleotides, BASB081

CC expression vectors, host cells, and a process for producing a BASB081

CC polypeptide. Also provided are vaccine compositions comprising a

CC BASB081 polypeptide or polynucleotide, and optionally at least 1

CC other M. catarrhalis antigen. A method for diagnosing a M.

CC catarrhalis infection involves identifying a BASB081 polypeptide, or

CC an antibody that is immunospecific for it, in a sample. A

CC therapeutic composition useful in treating M. catarrhalis diseases

CC in humans comprises an antibody directed against a BASB081 protein.

CC The disease can be a bacterial infection, e.g. otitis media in

CC infants and children, pneumonia in elders, sinusitis, nosocomial

CC infections and invasive diseases, chronic otitis media with hearing

CC loss, fluid accumulation in the middle ear, auditive nerve damage,

CC delayed speech learning, upper respiratory tract infection, and

CC inflammation of the middle ear.

XX

SO Sequence 2670 BP; 757 A; 527 C; 632 G; 754 T; 0 other;

Alignment Scores:

Pred. No.: 2,61e-10 Length: 2670

Score: 219.00 Matches: 130

Percent Similarity: 34.58% Conservative: 74

Best Local Similarity: 22.03% Mismatches: 232

Query Match: 6.90% Indels: 154

DB: 21 Gaps: 18

US-09-857-669-2 (1-609) x AAAS0537 (1-2670)

Qy 42 ProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAsp 61

Db 1294 CCTGCTATTGAACCGTGTAGCTAACGGATGGGATATTAAATGATATTTCGCCCATCGAA 1353

Qy 62 SerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGlu 81

Db 1354 TTTAGTCATCTCTAATCTGATTCAGACAGCTAAATTTGGTGGCTGCCAAGGCTGGCCAT 1413

Qy 82 ValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThr 101

Db 1414 TTATATCATCATGCTGATGATAGGTGCTTGCCATCAATCATGATGATGATGATGATG 1473

Qy 102 MetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAla 121

Db 1474 TCTATT-----TTGGCAGAAATCAGCGATGCGGTATCTGCC----- 1509

Qy 122 TyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIle 141

Db 1510 -----GTTGCAGGTGCTATT 1524

Qy 142 LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161

Db 1525 TTACCTGATGAATCTGAAATAGAGTAATAGATTTCGCCGAGCGTACCGCATTTGGCTAAT 1584

Qy 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181

Db 1585 CGCAAGACCCCTGCT-----GATGCTCATCAAGATAAAAA----- 1620

Qy 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAla 201

Db 1620 ----- 1620

Qy 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221

Db 1621 -----GTGCCGCTATATGTTCTTTGTGGCGAGTGATAAACCC 1656

Qy 222 -----IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237

CC and children, pneumonia in elderlies, sinusitis, nosocomial
 CC infections and invasive diseases, chronic otitis media with hearing
 CC loss, fluid accumulation in the middle ear, auditive nerve damage,
 CC delayed speech learning, upper respiratory tract infection, and
 CC inflammation of the middle ear.

XX Sequence 2760 BP: 775 A: 545 C: 654 G: 786 T: 0 other;

Alignment Scores:

Pred. No.: 2,73e-10 Length: 2760
 Score: 219.00 Matches: 130
 Percent Similarity: 34.58% Conservative: 74
 Best Local Similarity: 22.03% Mismatches: 232
 Query Match: 6.90% Indels: 154
 DB: 21 Gaps: 18

US-09-857-669-2 (1-609) x AAS0536 (1-2760)

QY 42 ProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAsp 61
 DB 1384 CCTGTCATTGAACCGTTGAGCTAACGGATGGGATATTAAATGGATATTTTCGCCCATCGAA 1443
 QY 62 SerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGlu 81
 DB 1444 TTTAGTGCATCTAATCTGATTCAAGACAAGCTAAATTTGGTGGCTCCCAAGGCTGCCAT 1503
 QY 82 ValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThr 101
 DB 1504 TTATATGACATGCCGTGATGATAGGCTTGGCCATCAATCATGATGATGGCTAAATCGC 1563
 QY 102 MetLeuArgSerLysGlyTyrPheSerSerLysValSerIleuThrGluLysAspGlyAla 121
 DB 1564 TCTAAT-----TTGGGCAAGATCAGCGATCGCTATCTGCC----- 1599
 QY 122 TyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIle 141
 DB 1600 -----GTTGCACGTGCTATT 1614
 QY 142 LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161
 DB 1615 TTACCTGATGAATCGAAATGAGGTAATAGATTGCCGAGCGTACCGCATTTGGCTAAT 1674
 QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
 DB 1675 CCAAGACCCCTGCT-----GATGCTCATCAAAATAAATA----- 1710
 QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAla 201
 DB 1710 ----- 1710
 QY 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
 DB 1711 -----GTCCCGCTATATGCTTTTGTGCGGAGTGATAAACC 1746
 QY 222 -----IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 DB 1747 CGAGATGGTCAAATTTGGCTGGGATCGGACACAGTACCGC----- 1794
 QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
 DB 1795 -----CTAGTC 1800
 QY 258 LeuAspPheGlnGlnAlaLeu---GluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
 DB 1801 ACAAAATTTGACCAATTTGATTAATCGTATGCTATCAACGAGCGCTGAGCTA--- 1857
 QY 277 AlaAspPheAspArgLeuGlnGlnGlyAspArgValProValLysValSerValThrGluVal 296
 DB 1858 -----AGACTGCTGAGGATAAAAGGGGTCAAGTTATATGCAACCAACCG 1905
 QY 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLys 316
 DB 1906 CTTAGCCACCTCTAATGATCAGCTAAGCAACT-----TTGGGT----- 1947

QY 317 IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
 DB 1948 -----TATCAACAAGAAGTTTTTGGTCACTCTACCAATGGT 1983
 QY 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
 DB 1984 TTTGATTTATCACACGCCCTAGAGCATGAGTTAGCCGAGTATTATCCAAATGGT 2043
 QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
 DB 2044 GGCTGG-----AATCGTACTATTTCATTCGCTTATCGTCTTGAAGCTTAAACCC 2094
 QY 377 AlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
 DB 2095 CAAGCACCCCTGAACATGCCAGGATTTACCAGTGGATTTTGTCAATGGTAAGCAAGC 2154
 QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
 DB 2155 CAAGAGCGGCTTATTGGCAGGTGTTGCTGTG----- 2184
 QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
 DB 2185 CATAAACG-----GTTGCAGATAAATTTGGTTAATCCG 2217
 QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
 DB 2218 ATGCGTGGCTAT-----CGTCAGCGATATTTCTTTAGAGGTGGCTCA 2259
 QY 457 ThrAlaLeuIleArgThrSer-----AlaArgAlaGly----- 467
 DB 2260 AGCGGTTTGGTATCGATGCTAATATGCTATTCGTCGAGCTGTTATTAGTGGCGTGTAT 2319
 QY 468 -----TyrPhePheThrProGluAsnLysLysLeuGlyThrPheIle 481
 DB 2320 AGTTTGGGGATAATGCTTATGGCAGCAATCGTCCCATCATGATGATGTTGGCATA--- 2376
 QY 482 IleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeu 501
 DB 2377 -----CAAGCAGGATACATTTGGTCGATAATTTTAATCATGTGCCATATCGTTG 2427
 QY 502 MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu 521
 DB 2428 CGTTTTTTCGTTGGTGGGACCAAAAGTATTCGTGGATATGCACATGATTTATACCT 2487
 QY 522 AlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr 541
 DB 2488 ATATCAGATAAGGTTATCTGACAGCGGTCAAGTATTGGCGGTGGTACAGCTGAATAT 2547
 QY 542 GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAla 561
 DB 2548 AATTATGAATTTATGAAGATTTGCGTTTGGCGGTTTTTGGTGATATTGTAATGCTTAT 2607
 QY 562 AlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
 DB 2608 GATAAAGGCTTTACTAATGATACCAAAATTCGTCAGGTGTCGTGTCCTGGCATCA 2667
 QY 582 ProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys-----LysIle 599
 DB 2668 CTTGTCGGTCAAGTTGCTGATGTGGCAACTGGTGTCAAGAGAGGCAATCCCAT 2727
 QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
 DB 2728 AAGCTCATTTTTTTTATTTGGCACACCATTT 2757
 RESULT 30
 AAF28347
 ID AAF28547 standard; DNA; 89047 BP.
 XX
 AC AAF28547;
 XX
 DT 04-APR-2001 (first entry)
 XX

DE Genomic fragment #34.
 XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW Bronchopulmonary; endocarditis; meningitis; ss.
 XX Moraxella catarrhalis.
 OS WO200078968-A2.
 XX 28-DEC-2000.
 XX 16-JUN-2000; 2000WO-US16649.
 XX 18-JUN-1999; 99US-0140121.
 XX (INCYTE GENOMICS INC.
 XX Lagace RE, Patterson C, Berg KL;
 XX WPI; 2001-041427/05.
 XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 XX Claim 1; Page 324-345; 545pp; English.
 XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;
 Alignment Scores:
 Pred. No.: 3,15e-08 Length: 89047
 Score: 219.00 Matches: 130
 Percent Similarity: 34.58% Conservative: 74
 Best Local Similarity: 22.03% Mismatches: 232
 Query Match: 6.90% Indels: 154
 DB: 22 Gaps: 18
 US-09-857-669-2 (1-609) x AAF28547 (1-89047)
 QY 42 ProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAsp 61
 DB 81780 CCTGTCATTGAACCGTTCAGCTACGGATGGGATATAATGGATATTTGCGCCATCGAA 81839
 QY 62 SerGluIleLysAspMetValGluHisLeuProLeuIleThrGlnGlnGluGlu 81
 DB 81840 TTTAGTGCATCTAAATCTGATTCAAGACAAGCTAAATTTGGTGGCTGCCAAGGCTGCCAT 81899
 QY 82 ValLeuAspLysGluGlnThrClyPheLeuAlaGluAlaProAspAsnValLysThr 101
 DB 81900 TTATATGATCCCTGATGATAGGGTGTGCCATCAATCATGATGCGGTAAATCCG 81959
 QY 102 MetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAla 121
 DB 81960 TCTATT-----TTGGCAGAAATCAGCGATGCCGTATCTGCC----- 81995
 QY 122 TyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIle 141
 DB 81996 -----GTTGACGTGCTATT 82010
 QY 142 LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn 161
 DB 82773 -----CAAGCAGGATACATTTGGTGGGATAATTTTAAATCATGTGCCATATCGTTTG 82823

DB 82011 TTACCTGATGAATCTCAAAATGAGGTAATAGATTGCCCCGAGCGTACCGGATTGGCTAAT 82070
 QY 162 TrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
 DB 82071 CGCAAGACCCCTGCT-----GATGCTATATCAAGTAAAAA----- 82106
 QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAla 201
 DB 82106 ----- 82106
 QY 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
 DB 82107 -----CTGCCGTATATGCTTTTGGCGAGTGTATAACCA 82142
 QY 222 -----IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
 DB 82143 CGAGATGGTCAAAATTTGGTGGGATGGGACACAGTACCCGC----- 82190
 QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
 DB 82191 -----CTAGTC 82196
 QY 258 LeuAspPheGlnGlnAlaLeu---GluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
 DB 82197 ACAAAATTTGACCATAAATTTGATTAAATCGTATGCTATCAAGCAGCGCTAGCTA--- 82253
 QY 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
 DB 82254 -----AGACTGCTCGAGATAAAAGGGTCAAGTTATATGCCACCAACCG 82301
 QY 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLys 316
 DB 82302 CTTAGCCACCCCTCTAAATGATCAGCTAAGCAACT-----TTGGCT----- 82343
 QY 317 IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
 DB 82344 -----TATCAACAAGAAGTTTGGTCTACTCTACCAATGGT 82379
 QY 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
 DB 82380 TTTGATTTATCCACACGCGACCCCTAGCATCAGATTAGCCGCAGTATTATCCAAATGGT 82439
 QY 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
 DB 82440 GCGTGG-----AATCGTACTTATTTCATGCGTATCGCTTTGATAAGTTAAACG 82490
 QY 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
 DB 82491 CAAGCACCCCTGAAACATGGCAGGATTTACAGTGGATTTTGTCAATGTTAAGCCAAGC 82550
 QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
 DB 82551 CAAGAGCGGTATTGGCAGGTGTGCTGTG----- 82580
 QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
 DB 82581 CATAAACG-----GTTGCAGATAATTTGGTTAATCCG 82613
 QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
 DB 82614 ATCCGTGGCTAT-----CGTCAGCGATATTCTTTAGAGGTTGGCTCA 82655
 QY 457 ThrAlaLeuIleArgThrSer-----AlaArgAlaGly----- 467
 DB 82656 AGCGGTTTGGTATCGGATCTAATATGGCTATTGCTCGAGTGGTATTAGTGGCTGTAT 82715
 QY 468 -----TyrPheThrProGluAsnLysLysLeuGlyThrPheIle 481
 DB 82716 AGTTTGGGGAATAAGTCTTATGGCAGCAANTCGTCCCATCATGACTGTGGGCATA--- 82772
 QY 482 IleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeu 501
 DB 82773 -----CAAGCAGGATACATTTGGTGGGATAATTTTAAATCATGTGCCATATCGTTTG 82823

QY 502 MetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeu 521
 Db 82824 CGTCTTTTGTGGTGGCGGACCAAGATATTCTGGATATGCATGACAGCTTATCACCCT 82883
 QY 522 AlaGlyProAsnGlySerValLeuProGluAlaAlaLeuValGlySerLeuGluTyr 541
 Db 82884 ATATCAGATAAGGTTATTCAGACAGCGGTCAAGATATTGGCGTGGTACAGCTCAATAT 82943
 QY 542 GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAla 561
 Db 82944 AATTATGAATTTTCAAGAAGATTGCGTGGCGGTTTTGGTGATATTGGTAATGCTTAT 83003
 QY 562 AlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSer 581
 Db 83004 GATAAAGGCTTTACTAATGATACCAAAATTTGGTGCAGGTGTCGGTTCGCTGGGCATCA 83063
 QY 582 ProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys-----LysIle 599
 Db 83064 CCTGTCGGTCAAGTTCGTGTGATGTGGCAACTGGTGTCAAGAAGAGGGCAATCCCATTT 83123
 QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
 Db 83124 AAGTGCATTTTATTTATTTGGCACACCATTT 83153
 RESULT 31
 ID AAQ66202 standard; DNA; 2989 BP.
 AC AAQ66202;
 DT 08-DEC-1994 (first entry)
 DE H. influenzae PAK 12085 D15 sequence.
 KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; NTHi; ss.
 OS Haemophilus influenzae PAK 12085.
 FH Key Location/Qualifiers
 FT CDS 389..2771
 FT /*tag= a
 XX
 PN W09412641-A.
 XX
 PD 09-JUN-1994.
 XX
 PF 23-NOV-1993; 93WO-CA00501.
 XX
 PR 23-NOV-1992; 92GB-0024584.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Chong P, Klein W, Loosmore S, Sia DYC, Thomas W;
 PI Yang Y;
 XX
 DR WPI: 1994-200269/24.
 DR P-PSDB; AAR53758.
 XX
 PT Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 PS Disclosure: Fig. 1E; 161pp; English.
 XX
 CC Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AA066198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC (AAR53754-58).

CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 2989 BP; 975 A; 468 C; 615 G; 931 T; 0 other;
 Alignment Scores:
 Pred. No.: 9,91e-10 Length: 2989
 Score: 213.50 Matches: 127
 Percent Similarity: 35.04% Conservative: 85
 Best Local Similarity: 20.99% Mismatches: 254
 Query Match: 6.73% Indels: 139
 DB: 15 Gaps: 24

US-09-857-669-2 (1-609) x AAQ66202 (1-2989)

QY 105 SerLysGlyTyrPheSerSerLysValSerLeuThr-----GluLys 118
 Db 1071 AATAATGGCTATGCCAAGCACAATACTTAAACGGATGTTTCAGCTAATGATGAAAA 1130
 QY 119 AspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys----- 133
 Db 1131 ACAAAAGTTAATGTAACCATTTGATGTAATCAAGGTTTACAGTATGACCTTCGTAGTGA 1190
 QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGly 149
 Db 1191 CGCATTATAGGTAATCTGGGAGGTATGCTGCCGAGCTTGAACCTTTTACTTCACCATTA 1250
 QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp 169
 Db 1251 CATTTAAATGATACTTTCGCCGCTAGT----- 1277
 QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
 Db 1278 ---GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTGGG-----GAACGA 1325
 QY 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp----- 206
 Db 1326 GGTAC-----GGTAACACACACAGTAATCTGTACCTGATTTTGACGAT 1370
 QY 207 ---ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
 Db 1371 GCAATAAACAATAGCATACCTTTGTTGATGCTGGACGAGTTTACTGTTCG 1430
 QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
 Db 1431 CAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATAGTACTTTAGCTCAGGAAATGCGA 1490
 QY 246 PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu 265
 Db 1491 CAACAAGAGGAACTTGGTATTAATTCACATTTAGTTGAGTACGAAAAATTCGCTTAGT 1550
 QY 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly--- 284
 Db 1551 CGTACAGGTTTCTTCGAA-----ACAGTTGAAACCGAATTGATCCTATCAATGCTAGC 1604
 QY 285 ---AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
 Db 1605 AATGATGAAGTGGATGCTCTATATAAACTCAAGACGTAACACGGGTAGTACACTTT 1664
 QY 304 GlyIleArgLeuAspSerGlyTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsn 323
 Db 1665 GGTATTGGTTAGGTACAGAGAGGTGATGATCAATTAACAAAGTATTAAACAAGATAAT 1724
 QY 324 LeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThr 343
 Db 1725 TTCCTGGGACAGGGCGGCGAGTATAGTATAGCTGTACGAAAAAATGATTATGTCAGCAT 1784
 QY 344 LeuAlaAlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsn 357
 Db 1785 GTCAATTTGGGTATACCGAACCACTATTATTAAGATGGTGTAAAGTCTTGTGGTGAAT 1844
 QY 358 TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg 376
 Db 358 TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg 376

QY 512 ArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGlu 531
 Db 627 CGTGGTTACGGCTATAAATAATGGCCCTAAATAAATAAATAATGGGAAATTCGTGGTGGC 686
 QY 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGly 551
 Db 687 TCACGTTTCTTACAGGTTCTTTAGAAATATCAATATCAAGTTTATCCGAAATGGTGGTG 746
 QY 552 AlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeuLysHis 571
 Db 747 GCACTTTTCAGATAGTATAGCGCTAA-TATTACACAGCAAAAGAGCTCGGTAT 805
 QY 572 GlySerGlyLeuGlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAla 591
 Db 806 GCGCAGCGGTGTGTGGTGGCATCGCCAGTGGCGCGGATTAAATTTTATATTTGCC 865
 QY 592 TyrGlyHisSerAsp-----LysLysIleArgTyrHisIleSerLeuGlyThr 607
 Db 866 ACACCCATTCGGATAAGATAACACAAATAATTCATTTTACRTTGGACTTGGTACA 925
 RESULT 33
 AAQ66201
 ID AAQ66201 standard; DNA: 2974 BP.
 AC AAQ66201;
 XX
 XX 08-DEC-1994 (first entry)
 DE
 XX H. influenzae SB33 D15 sequence.
 DE
 XX Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
 KW OMP; outer membrane protein; NTHi; ss.
 XX
 XX Haemophilus influenzae SB33.
 OS
 XX Key Location/Qualifiers
 FH 386..2764
 FT CDS /*tag= a
 FT
 XX W09412641-A.
 PN
 XX 09-JUN-1994.
 PD
 XX 23-NOV-1993; 93WO-CA00501.
 PF
 XX 23-NOV-1992; 92GB-0024584.
 PR
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
 PI Yang Y;
 PI
 XX WPI; 1994-200269/24.
 DR P-PSDB; AARS3757.
 XX
 XX Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 XX Disclosure; Fig. 1D; 161bp; English.
 PS
 XX Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AA066198-202) and the corresponding aa sequences were derived
 CC (AARS3754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 XX Sequence 2974 BP; 971 A; 471 C; 615 G; 917 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 1.87e-09 Length: 2974
 Score: 210.50 Matches: 121
 Percent Similarity: 34.88% Conservative: 75
 Best Local Similarity: 21.53% Mismatches: 241
 Query Match: 6.63% Indels: 125
 DB: 15 Gaps: 21
 US-09-857-669-2 (1-609) x AAQ66201 (1-2974)
 QY 134 IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu 151
 Db 1193 ATAGTAAATCTGGGAGGTATGCTGCGAGCTTGAACCTTTACTTTTCAGCATTTACATTTA 1252
 QY 152 AlaGluTyrTyrArgAsnAlaLeuGluAsnTyrGlnProValGlySerAspPheAsp 171
 Db 1253 AATGACTATTTCGCGCTAGT-----GAT 1276
 QY 172 GlnAspSerTyrGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
 Db 1277 ATTCACATGTGAAATGCAATTAAGCAAACTTGGG-----GAACGAGGTTAC 1327
 QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
 Db 1328 -----GGTAACACACAGTAAATTCGTACCTGATTTTGCAGCATGCAAT 1372
 QY 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
 Db 1373 AAAACATTAGCGATAACCTTTGTGTGATGCTGGACGACGTTTAACCTGTTTCACCACTT 1432
 QY 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
 Db 1433 CGCTTTGAAGGAATACCGTTTCTGCTGATAGTACTTACGTCAGAAATGCGCCACAA 1492
 QY 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
 Db 1493 CAAGCACTTGGTATAATTCACAAATAGTTAGTAGTAAAGAAATTCGCTTAGATCGTACA 1552
 QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
 Db 1553 GGTTCCTTCGAA-----ACAGTTGAAACCGAATGTGATCTCATCAATGGTAGCAATG 1606
 QY 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
 Db 1607 GAAGTGGATGCTGATATAAAGTCAAGAACGTAACACGGGTAGTAGTCAACTTGGTATT 1666
 QY 306 ArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
 Db 1667 GGTACGGTACAGAGGTGTTATAGTTATCAAGCAAGTGTCAACACAGATAATTTCTTG 1726
 QY 326 AsnLysGlyTyrIleGlySerValValTyrPaspMetAspLysTyrGluThrThrLeuAla 345
 Db 1727 GGAACAGGGCGCGCAGTAAAGTATAGTGGTACGAAATAATGATATGGTACGAGTCAAT 1786
 QY 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTyr 359
 Db 1787 TTGGGTTTATACCGACCTATTTTACTAAAGATGGTGAAGTCTTGGTGAATTTTTC 1846
 QY 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArgAlaPhe 378
 Db 1847 TTTGAAACTACGATAACTCTAAAGTATACATCTCTAACTATATAAGCGTACGACTTAT 1906
 QY 379 SerGlyGlyIle-----TyrTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
 Db 1907 GGAACATAATGTTACTTTAGGTTTCCCTGTAAGTAAATAAATCACTCTATTATTCAGGATTA 1966
 QY 396 Gly-----AlaGluPheLeuAlaGlu----- 402
 Db 1967 GCCCATACCTATAAATAATAGTAACTTTGCTCTAGAAATATACCGTAATTTATATATT 2026
 QY 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
 Db 2027 CAATCAATGAATTTAAAGGTAATGGCATTAACAAACAAATGACTTTCAT----- 2074

```
QY 417 HLAaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 2075 -----TTTTCTTTGGTGGAACTATAACAGCCTTAATAGAGCGCTATTTCCTCA 2122
QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 2123 ACTAAGGG-----GTTAAAGCAAGCTTGGTGGAGGTTACAAATTCAGGTTCTCGAT 2176
QY 457 ThrAlaLeuLeuLeuThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 2177 ACAAACTACTACAACTAGTCAGATGTCACAGGTTCTTACCCATTAGACAGATCAGATCAC 2236
QY 477 LeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAspAlaAsp 496
Db 2237 CTCTGGGTGTATCTGCAAAAGCATCTGCAGGATATGCAAAATGGTTTTGGAACAAGCGT 2296
QY 497 ValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeu 516
Db 2297 TTACCGTTCTATCAAACTTATACAGCGGGTGGCATTTGGTTTCATTACGCGGTTTTCCTTAT 2356
QY 517 AspPheIleGlyLeuAlaGlyProAsn----- 525
Db 2357 GGTAGCAAT-----GGCCCTAAGCCAATTTATCAAGGTCAAAATTAATAATTAAT 2407
QY 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
Db 2408 AGATAAGTTCTGATGTGATGGTGAATGCAATCGCTACAGCTAGCCAGGTAATAT 2467
QY 543 LeuPro-----PheThrArgThr----- 548
Db 2468 GTGCCAACTCCATTTGTGAGTGATAAGAGTCAAAATACAGTCGCAACCTCCTATTTCCT 2527
QY 549 -----LeuSerGlyAla 552
Db 2528 GATCGCGCAAGTGTTCGAATACTAATGGAATCAGATAAAATGGATTAGAGCAAT 2587
QY 553 ValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeuLysHisGly 572
Db 2588 GTCTTTGAAGACATTACCGGATTATGGC-----AAATCAAGCGGTACTCGCGCCCTCT 2638
QY 573 SerGlyLeuGlyValArgTyrPheSerProLeuAlaProPheSerPheAspIleAla--- 591
Db 2639 ACAGGTGTGGATGCCAATGCCAATCTCTAGTGGACCACTGGTATTTCTATCTCTAAA 2698
QY 592 -----TyrGlyHisSerAspLysIleArgThrPheHisSerLeuGlyThr 607
Db 2699 CCAATTAATAAATATGAAATGATGATGTCCAA---CAGTTCCAAATTTAGTATTGGGGT 2755
QY 608 ArgPhe 609
Db 2756 TCCTTC 2761
RESULT 34
AAQ66199
ID AAQ66199 standard: DNA; 2984 BP.
XX
AC AAQ66199;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae b Eagen D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; Hib; ss.
XX
OS Haemophilus influenzae type b Eagen strain.
XX
FI Key
FT Location/Qualifiers
FT CDS
FT 374..2767
FT /*tag= a
XX
PN W0912641-A.
```

```
XX 09-JUN-1994.
PD
XX 23-NOV-1993; 93WO-CA00501.
PF
XX 23-NOV-1992; 92GB-0024584.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
DR WPI: 1994-200269/24.
DR P-PSDB; AAR53755.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
PS Disclosure: Fig. 1B: 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2984 BP; 960 A; 474 C; 619 G; 931 T; 0 other;

Alignment Scores:
Pred. No.: 2,33e-09 Length: 2984
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 15 Gaps: 24

US-09-857-669-2 (1-609) x AAQ66199 (1-2984)
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 923 GAATCTGTAGTACAGTACATTAACAAGAACAAATGGAATTTACAACTGATCTTGGTGG 982
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
Db 983 AAATTTATGGGAAATAAATTTGAAGTGGCAATTCGAGAAGATTTCAGTCAATTCTGT 1042
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1043 GATTATTATTAATAATATGCTATGCCAAAGCACAAATTAATAAAGCGATGTTTCAGCTA 1102
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProThrLys 133
Db 1103 AATGATGAAAAACAAAGTTAATGTAACCATTTGATGTAATGAAGTTTACAGTATGAC 1162
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1163 CTTTCGTAGTCAGCGCATTTATAGGTAATCTGGAGGTATGCTGCGAGGTTGAACCTTTA 1222
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
Db 1223 CTTTCAGCATTTACATTTAAATGATACTTTCGCCGTAGT----- 1261
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1262 -----GATATTGCGAGTGTAGAAATGCAATTAAGCAAAACCTTGA--- 1303
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
```

[illegible]

Qy	526	-----GlySerValLeuProGlu 531
Db	2378	TATGGTAATGGTAGTGGTACTGGTACTTTAAAGAAAGATAAGTCTCTGATGTGATGGTGGT 2437
Qy	532	ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro 544
Db	2438	AATGCAATCGCTACAGTAGCGAGAGTTAAATGTGCCAACTCCATTTGTGAGCGATAAG 2497
Qy	545	PhethrArgThrLeuSerGlyAlaValPheHisAspMet 557
Db	2498	AGCCMAANTACGGTCCGAACCTCCATTATTGTGTATGCGGCAAGTGTGTGGAATACTAAA 2557
Qy	558	-----GlyAspAlaAlaAlaAsnPheLysArgMet 567
Db	2558	TGGAATACAGATAAAATGGATTAGAGAGCGATGATTAAAAAAGATTGCCTGATTATGGC 2617
Qy	568	-----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAla 584
Db	2618	AAATCAAGCGGTATTCGGCCCTCTACAGTGTGGATTCCAATGGCAATCTCCTATTGGG 2677
Qy	585	ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db	2678	CCATGTGATTCTCTTATGCGCAACCAATTAATAATATGAAAATGATGATGCGAA--- 2734
Qy	600	ArgTyrHisIleSerLeuGlyThrArgPhe 609
Db	2735	CAGTTCCAATTAGTATGAGGTTCTTTC 2764
RESULT 35		
ID	AAO66200	standard; DNA; 2953 BP.
AC	AAO66200;	
DT	08-DEC-1994	(first entry)
DE	H. influenzae b Minn A D15 sequence.	
KW	Vaccine; passive immunization; vector; antiserum; diagnosis; D15;	
KW	OMP; outer membrane protein; Hib; ss.	
OS	Haemophilus influenzae type b Minn A strain.	
FT	Key	Location/Qualifiers
FT	CDS	334..2727
FT		/*tag= a
FN	W09412641-A.	
PD	09-JUN-1994.	
PF	23-NOV-1993; 93WO-CA00501.	
PR	23-NOV-1992; 92GB-0024584.	
PA	(CONN-) CONNAUGHT LAB LTD.	
PI	Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;	
PI	Yang Y;	
DR	WPI; 1994-200269/24.	
DR	P-PSDB; AAR53756.	
PT	Nucleic acid encoding D15 outer membrane protein - esp. of	
PT	Haemophilus influenzae, and related proteins, vectors, antisera	
PT	etc. useful in vaccines, for diagnosis and for passive	
PT	immunisation.	
PS	Disclosure; Fig. 1c; 16lpp; English.	
CC	Outer membrane protein (OMP) D15 genes were isolated by screening	
CC	chromosomal libraries of H. influenzae type b (Hib) strains Ca,	

Db 2638 CCATTGGTATTCTCTTATGCAACCAATTAAAAAATATGAAAAATGATGATGTCGAA--- 2694

QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609

Db 2695 CAGTTCCAATTAGTATGGAGGTCTTTC 2724

RESULT 36

AAQ66198

ID AAQ66198 standard; DNA: 2949 BP.

XX AC AAQ66198;

XX DT 08-DEC-1994 (first entry)

XX DE H. influenzae b Ca D15 sequence.

XX KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;

XX KW OMP; outer membrane protein; Hib; ss.

XX OS Haemophilus influenzae type b Ca strain.

XX FT Key

FT CDS Location/Qualifiers

FT FT 75..2468

FT FT /*tag= a

XX WO9412641-A.

XX PN 09-JUN-1994.

XX PD 23-NOV-1993; 93WO-CA00501.

XX PR 23-NOV-1992; 92GB-0024584.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;

XX PI Yang Y;

XX DR WPI; 1994-200269/24.

XX DR P-PSDB; AAR53754.

XX PT Nucleic acid encoding D15 outer membrane protein - esp. of

XX PT Haemophilus influenzae, and related proteins, vectors, antisera

XX PT etc. useful in vaccines, for diagnosis and for passive

XX PT immunisation.

XX PS Disclosure: Fig. 1A; 16lpp; English.

XX CC Outer membrane protein (OMP) D15 genes were isolated by screening

XX CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,

XX CC Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK

XX CC 12085. Nucleotide sequences were determined for the D15 genes

XX CC (AAQ66198-202) and the corresponding aa sequences were derived

XX CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,

XX CC free of other antigens and lipooligosaccharides, by recombinant DNA

XX CC methods using the isolated genes.

XX SQ Sequence 2949 BP; 983 A; 477 C; 612 G; 877 T; 0 other;

Alignment Scores:

Pred. No.: 3.52e-09 Length: 2949

Score: 207.50 Matches: 130

Percent Similarity: 35.85% Conservative: 103

Best Local Similarity: 20.00% Mismatches: 280

Query Match: 6.54% Indels: 137

DB: 15 Gaps: 24

US-09-857-669-2 (1-609) x AAQ66198 (1-2949)

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80

Db 624 GAATCTGTAGTACGATACATTAACAAGAACAAATGGAATTACACCTGATTCCTGGTG 683

QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnVallys 100
 Db 684 AAATTTATGGGGAATAAATTTGAAGGTGGCAATTCGAGAAAGATTTGCAGTCAATGGT 743
 QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr 116
 Db 744 GATTATTATTTAAATAATGCTATGCCAAAGCACAAATTTACTAAACGCGATGTTCCAGCTA 803
 QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
 Db 804 AATGATGAAAAAACAAGTTAATGTAAACCAATGATTAATCAAGGTTTACAGTATGAC 863
 QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
 Db 864 CTTCTAGTCACGCATATATAGGTAATCTGGGAGGTATGCTGCCGAGCTTGAACCTTTA 923
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrAcuAsnAlaLeuGluAsnTrpClnGlnPro 165
 Db 924 CTTTCAGCATTTACATTTAATGATACTTTCCGCCGTAGT----- 962
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
 Db 963 -----GATATTGCAGATGTAGAAAAATGCAATTTAAAGCAAAACTTGGAA--- 1004
 QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 Db 1005 -----GAACGCGGTTACGGTAGCGCAACGCGTAATTCAGTA-----CCT 1043
 QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
 Db 1044 GATTTTGATGATGCAAAATAAACATTAGCGATAACCCCTGTTGTTGTTGCTGCGACGACGT 1103
 QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
 Db 1104 TTAAGTGTCCCAACTTCGCTTTGAAGAAATACCGTTCTGCTGATAGCAGCTTACGT 1163
 QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
 Db 1164 CAGGAAATGCGCCCAACAAGAGGAAGTGTGTAATAATTCACAATTTAGTTGAGTAGGAAAA 1223
 QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
 Db 1224 ATTGCTTAGATCGTACAGGTTCTTCGAA-----ACAGTCGAAACCGGAATGATCCT 1277
 QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
 Db 1278 ATCAATGCTAGTAATGATGAAGTGGATGTCGTATATAAAGTCAAAAGAACGTAACCGGT 1337
 QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
 Db 1338 AGTATCAACTTTGGTATGTTGTTACGGTACAGAGAGTGTATTAGTATTACAAGCAAGTCTT 1397
 QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
 Db 1398 AAACAAGATAATTTCTTGGGAACAGCGGCGCAGTAAGTATAGCTGGTACGAAATAATCAT 1457
 QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
 Db 1458 TATGTCAGAGTGTCAATTTGGGTTATACCGAGCCCTATTATTACTAAAGATGGGTAAAGT 1517
 QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
 Db 1518 CTTGCTGGAATGTTTCTTTGAAAACTAGCATAACTCTAAAGATGATACATCCCTCAAC 1577
 QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
 Db 1578 TATAAGCGTACGACTTACGGAAGTAATGTTTACTTTTAGGTTTCCCTGTAATAAATAATAC 1637
 QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
 Db 1638 TCCTATTATGTAGGATTAGGTCATACCTATAATAAAATAGTAACCTTTCCTCACAATAT 1697
 QY 403 -----GlyArgLysIleProGlySerAsp 410

```
Db 1698 AACCGTAATTTATATATTCATCAATCAATTAAGGTAATGGCATTAACAACAAATGAC 1757
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpIysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTTCTTTTGGTGGAACTATAACAGCCTT 1793
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 ATAGAGGCTATTCCCAACTAAAGG-----GTTAAACCAAGTCTTGGTGGACAGATT 1847
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCCAGGTCTCTGATAACAAATACTACAAACTTATACAGCGGTGGCATCGGTTCA 1907
Qy 471 ThrProGluAsnLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACCTCGGTTGTATCTGCAAAAGCATCTCCAGGATATGCAAT 1967
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTTTTGGAAACAAGCGTTTACCGTTCTATCAAACTTATACAGCGGTGGCATCGGTTCA 2027
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGGTTTCTTATGTAGTATT-----GGACCTAACCAATTTATGCCGAA 2078
Qy 526 -----GlySerValLeuProGlu 531
Db 2079 TATGCTAATGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2138
Qy 532 ArgAlaLeuLeuValGlySerLeuGlyTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCCAGAGTTAATGTGCCAAGTCCATTTCTGAGCGATAG 2198
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2199 AGCCAAATACGTCGCGCAACCTCTTATTTGTGATGGCGCAAGTGTGGTAATACTAAA 2258
Qy 558 -----GlyAspAlaAlaAlaAsnPhelysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGGATTAGACAGCGCATCTATTAAAGATTGCTCATTTATGGC 2318
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AAATCAAGCCGTATTTCGCGCTCTACAGGTGTCGGATTCCCAATGCGCAATCTCTATTGGG 2378
Qy 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGGTATTCTCTTATGCCAAACCAATTAATAAATATGAAATGATGATGCGAA--- 2435
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCTCAATTTAGTATTGGAGGTTCTTTC 2465

RESULT 37
AAF61065
ID AAF61065 standard; DNA; 2019 BP.
XX
AC AAF61065;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF08500.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.

XX PF 27-JUL-1999; 99DE-1035088.
XX PR 27-JUL-1999; 99DE-1035088.
XX PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX WPI: 2001-192469/20.
DR
XX New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria -
XX
PS Claim 1a: Page 117-118; 158pp; German.
XX
CC This invention describes novel DNA sequences (I) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (I); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (I) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (I) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
CC carrying one or more (I), (1), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (1),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 2019 BP; 404 A; 671 C; 603 G; 338 T; 3 other:

Alignment Scores:
Prod. No.: 7.59e-09 Length: 2019
Score: 201.50 Matches: 110
Percent Similarity: 34.36% Conservativeness: 57
Best Local Similarity: 22.63% Mismatches: 224
Query Match: 6.35% Indels: 95
DB: 22 Gaps: 19

US-09-857-669-2 (1-609) x AAF61065 (1-2019)
Qy 180 ThrSerValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAlaLys 195
Db 70 ACCACACGTCGACGCTGATCCCGCGCGCTGGGTACGAGGCTACACCTTCGCTAAC 129
Qy 196 LeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 215
Db 130 GTC---AACGGCGTGCACCAACCAACGACGACGACGACGACGACGACGACGACGACG 186
Qy 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlnArgTyr 235
Db 187 CTCGACCGCGGCAAGCGTCCCTACGTCAACGGCATCACTACCGCGGCAACGACGAC 246
Qy 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 247 GAAGACGAAGTGTGCGCGGAAATGCGGCASATGCGGAAAGCGCGCTGCGGCGACCTAC 306
Qy 256 LeuLeuAspPheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerVal 275
```

```
Db 307 CTGATCGACAGTCTCAGACCGCTGTGAGCGCTGGCTTCTTCAAGAGTCAACGTC 366
QY 276 GlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGlu 295
Db 367 GAGACCGCGCAGGTGCGTCGACGACGAGTCAAGTCAACGTACAGCGGTGGAAGAG 426
QY 296 ValLysArgHisLysLeuGluThrGlyLeuArgLeuAspSerGluTyrGly-----Leu 313
Db 427, CAGGCTCGCGCTCGATCAGCGCGCGCGGTTCGCCAGAGCGCGGCTGATCCTG 486
QY 314 GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerVal 333
Db 487 GGTGCTTCGATCAGCCAGAGC-----AACTTCCTCGGTACCGGTAAACAGGTATCCATC 540
QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyLeuSerGlnProArgAsn 353
Db 541 GCGCTGACCGCTTCGGAATACAGACCAAGTACAACTTCGGCTTCGTGATGCC----- 594
QY 354 TyrArgGlyAsnTyrTrpThrSer-----AsnValSerTyrAsn 366
Db 595 -----TACTTCACCGCGCAGCGGTGTGAGCTGGCTTACAACTGTCTTACAAAC 642
QY 367 ArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyLeu-----TrpTyrVal 385
Db 643 ACCACCGACTACGGGATAC-----TACGAGATGGTGTTCGTACTATGCA 690
QY 386 ArgAspArgAlaGlyIleAspAlaArgLeuGly-----AlaGluPheLeu 400
Db 691 ATCAACAGCTACGCGCGCGCGGTGAGCTTTGGCTACCGATCAACGAAACCTCGCGCGCTG 750
QY 401 AlaGluGlyArgLysIleProGlySerAspPheAspLeuGlyAsnSerHisAlaThrMet 420
Db 751 ACCTACGCGCTGACCGCTGCAGCAGCAGACATTTCCCGCGGACCTACAGCGCTCAGCAG 810
QY 421 Leu-----ThrAlaSer 424
Db 811 ATCTACGATTTTCATCTCCCGGAAGCAAGAGCTTCAACAACCTCAAGCGCTCGATCGGT 870
QY 425 TrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAsp-- 443
Db 871 TGGTGGGAATCTACCTCAACAAGGCGTACTGCGCACCGCGGCGCTCGCAGACGCTG 930
QY 444 GlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
Db 931 AGCGTACGGTGTACACCGCTTCGCGAGCGAGCTGTCTTACAAACTCGACTACAAC--- 987
QY 464 AlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg 483
Db 988 -----GGCAGGCTTCTCCTCGCGGTGACCAGCAACACC---ACGCTGCGCGCTGCAC 1035
QY 484 GlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAsp---ValProSerGlyLeuMet 502
Db 1036 ACCAACCTGGGTGCGCAACGGCTATGGCTCCACCGCGCTTCGCTTCTATGAAAGC 1095
QY 503 PheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAla 522
Db 1096 TACACGCGCGTGGCGAGCGGTGCGTACGCGGGTTCAAG-----CACGGCACCTC 1146
QY 523 GlyProAsnGlySer----- 527
Db 1147 GCGCGCGCAGACACACCGGCTACCGGGCGCTATTCCAACGCTGGGAGGCTTACTATTCC 1206
QY 528 -----ValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr--- 541
Db 1207 GACCGCGACACCGATGCTCGGGGGGCANTATCTGATTAACGCGCGCTCCGAGTACCTG 1266
QY 542 ---GlnLeuProPheThrArg-----ThrLeuSerGlyAlaValPheHisAspMet 557
Db 1267 TTCCCGCTGCCCTTCGTGAAGGACGACGAGCAACTGGCGGACGTCGTGTTCGTATGCC 1326
QY 558 GlyAspAlaAlaAsnPheLysArgMetLysLeuLysHisGlySer----- 573
```

```
Db 1327 GGCACAGTGTATCGGACACCTGCTACCTGTCCACCAAGGCTGCGGACGCTCGAC 1386
QY 574 -----GlyLeuGlyValArgTrpPheSerProLeuAlaPro 585
Db 1387 CTGGCGCAGATGGCTGTGCTGGGGTTGGCGTACCTGGTACAGCCGATGGGCGC 1446
QY 586 PheSerPheAspIleAla 591
Db 1447 CTGAGTTTCACGCTGGCG 1464
RESULT 38
AAV25138
ID AAV25138 standard; DNA; 2778 BP.
XX
AC AAV25138;
DT
XX
XX 13-JUL-1998 (first entry)
DE H. pylori cell envelope OMP ORF 05ep10815_4719175_cl_115.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW cell envelope outer membrane protein; OMP; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..2778
FT FT /*tag= a
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
XX (ASTR ) ASTRA AB.
XX
XX Alm RA, Smith D.
XX
XX WPI: 1997-503122/46.
XX P-PSDB, AAW55729.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
XX infection and for diagnosis of H. pylori infection
XX
XX Claims 5,6,21; Page 929,930; 1145pp; English.
XX
XX This sequence encodes a H. pylori cell envelope outer membrane
XX protein having a terminal phe residue.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds,
XX useful as potential H. pylori life cycle activators or inhibitors. The
XX DNA and probes derived from it may be used for the identification of
XX H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
XX acid sequences complementary to the DNA act as antisense sequences and
XX can be used to prevent the translation of H. pylori mRNA. Antibodies
XX against the protein can be used in immunoassays to evaluate the abundance
XX and distribution of H. pylori-specific antigens. The genomic sequence of
XX H. pylori (ATCC 55679) was determined from overlapping contigs generated
XX by mechanically shearing the bacterial DNA. The sequences were analysed
XX for ORF of at least 180 nucleotides, and the predicted coding regions
XX defined by computer evaluation. To identify likely H. pylori antigens for
XX vaccine development, the amino acid sequences predicted from various ORF
XX were analysed for significant homology to other known or exported
```

CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from *H. pylori* by PCR
CC amplification for recombinant polypeptide production, e.g. in *E. coli*
CC hosts.

XX Sequence 2778 BP; 813 A; 514 C; 681 G; 770 T; 0 other;

Alignment Scores:

Pred. No.: 4, 25e-06 Length: 2778
Score: 174.00 Matches: 126
Percent Similarity: 33.67% Conservatism: 72
Best Local Similarity: 21.43% Mismatches: 197
Query Match: 5.48% Indels: 194
DB: 18 Gaps: 23

US-09-857-669-2 (1-609) x AAV25138 (1-2778)

```
Qy 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
   ::::::::::::::::::::
Db 1102 ATGCGCGATAGGGCTATCGCTT-----CGGTGGTGAAGCCA 1140

Qy 206 AspThrAlaThrValAspLeuAsnValVal-----AspSerGly 219
   :::::
Db 1141 GACTTGGATAAAGACGAAAAACGGGCTTCTGAAAGTCATTTATCGTATTGAAGTGGGC 1200

Qy 220 ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle 239
   ::::::::::::::::::::
Db 1201 GATATGGTCATATCATGTCATATTCAGGGAACCGACGACGACGACGACGACGACGAC 1260

Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
   :::::
Db 1261 ATTAGGAGGAATTTACTAGGGCTTAAGATAAATACAACTTGACCAACTGAGAAT 1320

Qy 260 PheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
   ::::::::::::::::::::
Db 1321 TCGAAATTTCTTGGCGGTTAGGGTTAGGTTTCTCT-----AAAGTCAAGATTGAAGAA 1374

Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
   :::::
Db 1375 AAAAGGTCATATAGCTCATTTGATGATTTGTTAGTAGCGTAGAAGCGGCGCACCGGG 1434

Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
   ::::::::::::::::::::
Db 1435 CACTTGCATTCGGGTGGGTGGCTCTATGAGGCGCTCATGTTAATGGGACCGTG 1494

Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp----- 336
   ::::::::::::::::::::
Db 1495 AGCGAAGGAATCTTTTGGCACAGGGCAAGCATGAGCTTGATGCTAACATTGCCACA 1554

Qy 337 -----MetAspLysTyrGluThrThrLeuAlaLaGly--- 347
   :::::
Db 1555 GGGGGGGTAGATCTATCCGGGATCCCAAGGGCGGGCGGTATGTTGCCGGGAAT 1614

Qy 348 -----IleSerGlnProArg----- 352
   :::::
Db 1615 TTGAGCTTGACTAATCCCAAGGATTTTTCACAGCTGGTATAGCTCTACGATCAATCTTTAT 1674

Qy 353 ---AspTyrArg----- 355
   :::::
Db 1675 CGCGATTACAGGATAGCTACCAATACATCCAAAGCGGGGGCTTTGGGGTGAATGTC 1734

Qy 356 -----GlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln 371
   ::::::::::::::::::::
Db 1735 GGGCGCATGCTGGGTAAATAGAACCCCATGTGAGCTTAGGTATTAAGTGAATGTACCAAA 1794

Qy 372 AsnLeuGluLysArgAlaPheSerGlyGlyIleThrTyrValArgAspArgAlaGlyIle 391
   :::::
Db 1795 CTCCTT-----CGTTTCAGGACCCCTTA---TACAAACCGCTACTATTCCTCTGTT 1842

Qy 392 Asp-----AlaArgLeuGlyAlaGluPheLeu 400
   :::::
Db 1843 AATGAAGTGGGCTCTCCAGGCAATGTTCCACACCCGCAATCGGTGATTATCAACCGCTTA 1902
```

```
Qy 401 AlaGluGlyArgLys-----IleProGlySerAspIleAspLeuGly----- 414
   ::::::::::::::::::::
Db 1903 TCAGGGGGTAGAACTCCCATTTGGTCTGAAAGCTGTCTTAGTCTCGAGGCGATCACCAC 1962

Qy 415 -----AsnSerHisAlaThrMetLeuThrAlaSerTyr 425
   :::::
Db 1963 TCACCAGAAATAAAGGTATTGGGATAGGGATTACCACACGCTATCACCAGTCTCTTC 2022

Qy 426 LysArgGlnLeu-----LeuAsnAsnValLeuHisProGluAsnGlyHisTyr 441
   :::::
Db 2023 ACCCTTGATGTGAGCTATGACACACCGATGATTATTTCCTCTAGAATGGGGTATC 2082

Qy 442 LeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu-Ile 461
   :::::
Db 2083 TTTAGTTCCTAT-----GCGACAATGCTGTTTGGTTCGAAGCTCTGGCAGCTCAATTCT 2136

Qy 461 gThrSerAlaArgAlaGlyTyrPhePheThrProGlu-----AsnLysLysLeuG 478
   :::::
Db 2137 TGGACGGGTTAGGCGGAATGTCCTGAACACCAAGATTTATGCTAAATTCGCCGCTTAC 2196

Qy 478 yThr-----PheIleIleArgGlyGlnAlaG 487
   :::::
Db 2197 CACCATTTGCAAAAAATATTATTATGATAGATTGATCGCTCTTTTAAACG-CAAGGGG 2255

Qy 487 yTyrThrValAlaArgAspAsnAlaAsp-----ValProSerGlyLeuMetPheArg 505
   ::::::::::::::::::::
Db 2256 CTAT---ATCTTAGGTATAACCGGATGATTACTTCCCTTAACTCCGCTTCTTACAT 2312

Qy 505 rGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAs 525
   ::::::::::::::::::::
Db 2313 GGGGGCGTAACACGCTGAGAGGCTTTAGG-----AA 2345

Qy 525 nGlySerValLeuProGlu----- 531
   ::::::::::::::::::::
Db 2346 CGGCTCAATCACACCTAAAGATGATTTGGCTTGTGGCTTGGAGGCGATGGGATTTTAC 2405

Qy 531 ----- 531
   ::::::::::::::::::::
Db 2406 CGCTTCTACTGAATTTGAGCTATGGGCTTAAACAGCGCTTAAATCGCTTACGGTGT 2465

Qy 532 -----ArgAlaLeuValGlySerLeuGlyTyrGlnLe 543
   :::::
Db 2466 TTTTGCCTTGGTTTCTTAACCTTTAAACCCCACTAGGGGAGTCTTCTTATTAACGC 2525

Qy 543 uProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAs 563
   :::::
Db 2526 TCCACACGACGAG-----GCCAATTTTAAAGATTATGGCGTTGTAGGGCGTGG 2573

Qy 563 nPheLysArgMetLysLeuLysHisGlySerGlySerGlyValArgTyrPheSerProLe 583
   :::::
Db 2574 GTTGTAAAGGGCGACTTGGAGGGCTTCTACAGGCTTACAGTTGAATGATTTTCGCCCAT 2633

Qy 583 uAlaPro-----PheSerPheAspIleAlaTyr----- 592
   :::::
Db 2634 GGGGGCTTTGGTGTGATTTTCCCTATAGCGTTTTCACACCAATGGGGCGATGGCAATGG 2693

Qy 593 -----GlyHisSerAspLysLysIleArgTrpHis 602
   :::::
Db 2694 CAAAAAATGAAAGGCTGTGCTTTTAAACCTTAACCTAAGATTAACGCAACATTTTGA 2753

Qy 602 sIleSerLeuGlyThrArgPhe 609
   ::::::::::::::::::::
Db 2754 ATTTTCTATGGGAACAAGGTTT 2775

RESULT 39
ID AAX75802
XX AAX75802 standard; DNA; 2778 BP.
AC AAX75802;
DT 03-AUG-1999 (first entry)
XX
DE H. pylori outer membrane polypeptide encoding DNA.
```


XX Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
 KW cellular immune response; ds.

XX Helicobacter pylori.

XX W09921959-A2.

XX 06-MAY-1999.

XX 28-OCT-1998; 98WO-US22883.

XX 17-DEC-1997; 97US-0993001.

PR 28-OCT-1997; 97US-0959131.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;

XX WPI; 1999-326698/27.

DR P-PSDB; AAY117183.

XX Cellular vaccine against Helicobacter pylori

XX Claim 8; Page 147-149; 352pp; English.

XX The invention relates to a vaccine for preventing or treating infections
 CC by Helicobacter pylori. The vaccine contains at least one isolated
 CC H. pylori polypeptide, or its fragments, in a carrier, where the
 CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
 CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
 CC humoral and cellular immune responses. The vaccines are used to treat or
 CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
 CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
 CC (OMPs) AAY17160 to AAY17218.

XX SQ Sequence 2778 BP; 813 A; 514 C; 681 G; 770 T; 0 other;

Alignment Scores:

Pred. No.: 4.25e-06 Length: 2778
 Score: 174.00 Matches: 126
 Percent Similarity: 33.67% Conservative: 72
 Best Local Similarity: 197 Mismatches: 194
 Query Match: 5.48% Indels: 23
 DB: 20 Gaps: 23

US-09-857-669-2 (1-609); x AAX75802 (1-2778)

QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205

DB 1102 ATCCCGGATAGGGCTATGCGTTT-----CGGTGGTGAAGCCA 1140

QY 206 AspThrAlaThrValAspLeuAsnValVal-----AspSerGly 219

DB 1141 GACTTGGATAAAGACGAAAAAGGGGCTGTGAAAGTCATTTATCGTATTGAAGTGGCC 1200

QY 220 ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle 239

DB 1201 CATATGGTGCATATCAATGATGTCATATTCAGGAAACCGGACGAGCGATAGGATC 1260

QY 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259

DB 1261 ATTAGAGGGAATTTCTTACTAGGCGCTAAAGATAAATACAACTTGACCAACTGAGAAAT 1320

QY 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279

DB 1321 TCGGAAATTCCTTGGCGGCTTATAGGTTTCTCT-----AAAGTCAAGATTTGAAGAA 1374

QY 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299

DB 1375 AAAAGGTCATAGTCATATGATGATTTGTTAGTACGCTAGAGAGGGGCGACCGGG 1434

QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319

DB 1435 CAGTTTGCATTTCCGGTTGGGCTATGGCTCTTATGGAGGGCTCATGCTTAATGGGAGCGTG 1494

QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp----- 336

DB 1495 AGCGAAGGAATCTTTTGGCAGAGGCAAAAGCATGAGCTTGTATGCTAACAATGCCACA 1554

QY 337 -----MetAspGlyTyrGluThrThrLeuAlaAlaGly--- 347

DB 1555 GGGGGGGGTAGATCTTATCCGGGCATGCCAAAAGGGGGGGGGGTATGTTCCCGGAAT 1614

QY 348 -----IleSerGlnProArg----- 352

DB 1615 TTGAGCTTGACTAATCCAAGGATTTTTCACAGCTGGTATAGCTCTACGATCAATCTTTAT 1674

QY 353 ---AsnTyrArg----- 355

DB 1675 GCGGATTACAGGATAAGCTACCAATACATCCAAAGGGGGGGGGCTTTGGGTGAATGTC 1734

QY 356 -----GlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln 371

DB 1735 GGGCGCATGCTGGGTAAATAGAACCATGTCAGCTTAGGGTATAAAGTGAATGTTACCAAA 1794

QY 372 AsnLeuGluLysArgAlaPheSerGlyIleTyrTyrValArgAspArgAlaGlyIle 391

DB 1795 CTCCCTT-----GGTTTCAGCAGCCCTTA---TACAACGCTACTATTCTCTCTGTT 1842

QY 392 Asp-----AlaArgLeuGlyAlaGluPheLeu 400

DB 1843 AATGAAGTGGCTCTCCAAAGCAATGTTCCACACCCGCAATCGGTGATTTCAACCGCTTA 1902

QY 401 AlaGluGlyArgLys-----IleProGlySerAspIleAspLeuGly----- 414

DB 1903 TCAGCGGTAGAGATCCCATGTTGCTTCTGAAAGCTGTTCTAGTCTCGGACGGATCAACCACT 1962

QY 415 -----AsnSerHisAlaThrMetLeuThrAlaSerTrp 425

DB 1963 TCACCAGAAATAAAGGTATTTGGGATAGGATTAACCAACAGCTATCAACAGCTCTTTC 2022

QY 426 LysArgGlnLeu-----LeuAsnAsnValLeuHisProGluAsnGlyHisTyr 441

DB 2023 ACCCTTGATGGAGCTATGACACACGATGATTATTTATTTCCCTAGAATGGGTATC 2082

QY 442 LeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu-IleArg 461

DB 2083 TTTAGTTCTCTAT-----GCGCAATGTCTGGTTTGCACAGCTCTGCCAGCTCAATCT 2136

QY 461 gThrSerAlaArgAlaGlyTyrPhePheThrProGlu-----AsnLysLysLeuG 478

DB 2137 TGSACGGGTTAGGCGGAATGTCCTGTAACACCAAGTTTATGTAATTCGCCGCTTAC 2196

QY 478 YThr-----PheIleArgGlyGlnAlaG 487

DB 2197 CACCATTTGCMAAATAATTTATTTGATAGATTTGATCGCTCGTTTAAAAACG-CAAGGGG 2255

QY 487 YThrThrValAlaArgAspAsnAlaAsp-----ValProSerGlyLeuMetPheArgSe 505

DB 2256 CTAT---ATCTTTAGTATACACCGATGATTACTTCCCTTAAACTCCACCTTCTACAT 2312

QY 505 rGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAs 525

DB 2313 GGGGGCGGTAAACACGGTGAGAGGCTTTAGG-----AA 2345

QY 525 nGlySerValLeuProGlu----- 531

DB 2346 CGGCTCAATCACACCTAAAGATGAGTTTGGCTTGGCTTGGAGGCGATGGGATTTTAC 2405

QY 531 ----- 531

DB 2406 CGCTTCTACTGAATTTGAGCTATGGGTGTTAAAAAGCGGCTTAAATGCGTTAGCGTGGTT 2465

QY 532 -----ArgAlaLeuLeuValGlySerLeuGlyGlnLe 543

Db 2466 TTTTGACTTTGCTTTTAACTTTAAACCCCACTAGGGGAGTTTCTTCTATAACGC 2525
QY 543 uProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAs 563
Db 2526 TCCACACGACG-----GGCAATTTTAAAGATTAATGCGGTGTAGGGGCGTGG 2573
QY 563 nPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLe 583
Db 2574 GTTTGAAGGCGGCTTGGAGGCTTCTACAGGCTTACAGATGAATGGATTTCGCCCAT 2633
QY 583 uAlaPro-----PheSerPheAspIleAlaTyr----- 592
Db 2634 GGGGCTTTGGTGTGATTTCCTATAGCGTTTTCACCAATGGGCGATGGCAATGG 2693
QY 593 -----GlyHisSerAspLysLysIleArgTrpPhe 602
Db 2694 CAAAAAATGTAAGGCGTGTCTTAAACCTTAACATGAACGATTACACCAACATTTGA 2753
QY 602 sIleSerLeuGlyThrArgPhe 609
Db 2754 ATTTCTATGGGAACAAGGTTT 2775
RESULT 40
ABL91257
ID ABL91257 standard; DNA; 2373 BP.
XX ABL91257;
AC
XX
DT 29-JUL-2002 (first entry)
XX
XX Chlamydia pneumoniae cp6576 ORF DNA, SEQ ID NO:148.
DE
XX Chlamydia pneumoniae; antigen; immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029; open reading frame; ORF; gene; ds.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1..2373
FT /tag= a
FT /product= "cp6576"
FT sig_peptide 1..78
FT /tag= b
FT mat_peptide 79..2370
FT /tag= c
FT /product= "Mature protein"
XX
PN WO200202606-A2.
XX
XX 10-JAN-2002.
XX
XX 03-JUL-2001; 2001WO-IB01445.
XX
XX 03-JUL-2000; 2000GB-0016363.
XX 11-JUL-2000; 2000GB-0017047.
XX 21-JUL-2000; 2000GB-0017983.
XX 07-AUG-2000; 2000GB-0019368.
XX 18-AUG-2000; 2000GB-0020440.
XX 14-SEP-2000; 2000GB-0022583.
XX 10-NOV-2000; 2000GB-0027549.
XX 22-DEC-2000; 2000GB-0031706.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Rattli G, Grandi G;
XX
XX WPI: 2002-154726/20.
DR N-PSDB; ABB90599.
XX
XX Novel Chlamydia pneumoniae protein useful in the manufacture of a

PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -
XX
PS Claim 5; Page 113-114; 364pp; English.
XX
CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
CC the invention.
XX
SQ Sequence 2373 BP; 764 A; 457 C; 448 G; 704 T; 0 other:
Alignment Scores:
Pred. No.: 3.82e-06 Length: 2373
Score: 173.50 Matches: 154
Percent Similarity: 34.40% Conservative: 93
Best Local Similarity: 21.45% Mismatches: 257
Query Match: 5.47% Indels: 215
DB: 24 Gaps: 34
US-09-857-669-2 (1-609) x ABL91257 (1-2373)
QY 22 ProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSer 41
Db 259 CCTAAAGTAGAATTTCTGAAGGAAA---ACTAAACATAGCCCTTCACCTAATAGTAAA 315
QY 42 ProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAsp 61
Db 316 CCCTCAATTCGAATATTCATATCTCAGGA----- 345
QY 62 SerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGln-----Gln 79
Db 346 -----AATCAAGTCGTTCCCTGAACATAAAATTTTAAACCCCTACAAATTTACCGT 396
QY 80 GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnVal 99
Db 397 AATGATCTCTTTTGAACGAGAAAA-----TTTCTT-----AAGGGTCTTGATGATCTA 444
QY 100 LysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAsp 119
Db 445 AGAACCTATTATCTCAAGCGAGGATATTTCCGATCCAGTACAGTACAGTCTGGAACAC 504
QY 120 GlyAlaTyrThrValHis-----IleThrProGlyProArgThrLys 133
Db 505 AATCAAGAAAAAGGTCACATCGATGTTTAAATTAATAATCAATGAAGTCTCTCGCGAAA 564
QY 134 IleAlaAsnValGlyValAlaIleLeuGly----- 143
Db 565 ATTAACACGCTTACGTTCTCAGGAATCTCTCGATCAGAAAAATCAGATATCCAGAATTT 624
QY 143 ----- 143
Db 625 ATTCAAACCAAGCAGCACTCTACAACCTACAAGTTGGTTTACTGGAGCTGGACTCTATCAC 684
QY 144 ---AspIleLeuSerAspGlyAsnLeuAla-----GlutryTyr 155
Db 685 CCAGATATTGTTGTAACAAGATAGCTTGCAATTTACGAAATTTACCTACATATACCGGGTAC 744
QY 156 ArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrp 175
:::||||::: ||| ||| ::||| |||

Db 745 GCTGATGCTATA-----GTCACACTCTCACTATACACCTTGACGACAAA 786
QY 176 GluAsn-----SerLysThrSerValLeuGlyAlaVal 186
Db 787 GGAATATTCTTTTACATGGATATTGATCGAGGTCGCGATATACCTTAGGACACGTC 846
QY 187 ThrArgLysGlyTyrProLeu----- 193
Db 847 CATATCCAGGGTTTGAGTTTGCCAAACGCCCTTATAGAAAAGCAATCCCAAGTCGCG 906
QY 193 ----- 193
Db 907 CCCAATGATCTTTATTGGCCCGATAAAATATGGGATGGGCTCATAGATCAAAACAACT 966
QY 194 ---AlaLysLeuGly-----AsnThrArgAlaAlaVal-----AsnProasp 206
Db 967 TATGCAAAAGTATGGCTACATCAATACCAATGCTAGACGCTTCTCTTCATCCCTCAGCAACC 1026
QY 207, ThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyasp 226
Db 1027 CGCCCTATTATTGATGTAACCTTATAGGTAAAGTGAAGGGCTCTCTTATAAGTTGGGTTA 1086
QY 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
Db 1087 ATTAAATTTACTGGGAATACCATACATAAAATCTGCAGCTTATTTTACACGAAACCACTCTC 1146
QY 247 GlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGln 266
Db 1147 TTCCCAAGAGATACATTCATCGCTTAAAGCTAGAAGATACCTGAGCAACGTTTAAAGAAAT 1206
QY 267 AsnGlyHisTyrSerGlyAlaSerVal-----GlnAla 277
Db 1207 ACAGGCTACTCCAAAGCGTTAGTCTCTATACAGTCTGCTTCAACTGATCCTATGGGC 1266
QY 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGlu----- 295
Db 1267 AATCGGATCAATACCGAGATATTTTGTAGAAGTCAAAAGAAAACAAACAGGAAACTTGA 1326
QY 296 -----ValLysArgHisLysLeuGluThrGlyIleArgLeu---Asp 308
Db 1327 GGCTTATTCTTAGGATTAGTCTCTTGACAACTCTTTTGGAGAAATGAACATCTCTGAA 1386
QY 309 SerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGly 328
Db 1387 AGTAATTTTGTATCTATTGGAGCT-----AGAAATATATTTTCTAAAGGT 1431
QY 329 Tyr-----IleGlySerVal 333
Db 1432 TTTCTGTTGCTAAGAGCGGTGGAGAACATCTATTCTTAAAGCCAACTTCGGGGACAAA 1491
QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
Db 1492 GTCACAGACTATACTTTGAAGTGGACC-----AAACCT---CAT 1527
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnVal-----SerTyrAsnArgSerThr 369
Db 1528 TTTCTAAACACTCTTGATTTTAGGAATTCAGATTAATCAATTAACAGAGCATTA 1587
QY 370 ThrGlnAsnLeuGluLysArgAlaPheSerGlyIle-----TrpTyrValArgAsp 387
Db 1588 TCTAAAGATTATGCTGTCCAAACCTATATGGCGGAAACGCTCAGCACAAACGATATCTTTGAAC 1647
QY 388 -----ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArg 404
Db 1648 GAACACCTGAATACGGTCTATTTTATCGAGGAAGTCAACAGGATTTACATGAAGAAACGT 1707
QY 405 Lys-----IleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
Db 1708 AAGTCTCTCTAGGCGCAATATAGAC-----AGCAATAAGGATTTTGTCTCTGCT 1758
QY 424 SerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAsp 443
Db 1759 GCAGGTGCTCAACTTGAATACGATTTCTGTAGATAGTCTCTAG-NACTCCAACTACAGGAT 1817

QY 444 GlyLysIleGlyThrThrLeuGlyThrPhe-LeuSerSerThrAlaLeuIleArgThrSe 463
Db 1818 TCCCGGGGGGGTGCACCTTTGAGGTTTCTGTTTGGGAGGAAC- 1860
QY 463 rAlaArgAlaGlyTyrPhePheThrProGluAsn-----LysLysLe 477
Db 1861 -----TATCATTTTACAAAACCTCTTTAAACAGCTCTATCTATAGAAAAC 1907
QY 477 u-----GlyThrPheIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAs 494
Db 1908 TACGCGTAAGGTATTTTGAATAATCAAGGGGAAGCTCAATTTATTAAACCCCTATAGCAA 1967
QY 494 nAla-----AspValProSerGlyLeuMetPheArgSerGlyAlaSerSerVa 511
Db 1968 TACTACAGCTGAAGCAGTTCCTGTCAGTGAGCGCTTCTTCTTAGTGGAGAGACTACACT 2027
QY 511 lArgGlyTyrGluLeuAspSerIleGly-----LeuAlaGlyProAsnGlySe 527
Db 2028 TCGGGGATATAAATCCTTTATATTCGGTCCAAATACTCTGCTACAGAAACCTCAGGGAGG 2087
QY 527 rValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrAr 547
Db 2088 ACTC-----TCTTCGCTCCTTATT---TCAGAAGAGTTCATACCTCTCATCAG 2135
QY 547 g-----ThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhely 565
Db 2136 ACAACCTAATATTAGTGCCTTTGATTCTTAGACTCAGGTTTTCGTTTACAAGAGTA 2195
QY 565 sArgMetLysLeuLys-----HisGlySerGlyLeuGlyValArgTrp 579
Db 2196 TAAGATTTTCGTTAAAGATCTACGTAGTAGTGTGCTGATTTGGTCTGCGCTTC 2247

Search completed: November 9, 2002, 01:37:16
Job time : 726 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:13:49 ; Search time 35 seconds
(without alignments)
2318.561 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMTKTPALLPALFFPPHAY.....IAYGSDKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3173	100.0	609 21	AAV99623 Neisseria meningitidis
2	3173	100.0	609 21	AAV99624 Neisseria meningitidis
3	3153	99.4	615 21	AAV74862 Neisseria meningitidis
4	3136	98.8	615 21	AAV74861 Neisseria meningitidis
5	3019	95.1	587 21	AAV99625 Neisseria meningitidis
6	1443	45.5	522 21	AAV74860 Haemophilus influenzae
7	374	11.8	578 21	AAV95820 Amino acid sequence
8	355.5	11.2	576 21	AAV95821 Neisseria meningitidis
9	260.5	8.2	792 21	AAV84946
10	256.5	8.1	792 22	AAU03958

11	256.5	8.1	792	22	AAV84745
12	256.5	8.1	792	22	AAU23786
13	253	8.0	797	22	AAU03957
14	253	8.0	797	22	AAU03959
15	253	8.0	797	22	AAU04451
16	253	8.0	797	22	AAU04451
17	253	8.0	797	22	AAU04451
18	253	8.0	797	22	AAU04451
19	253	8.0	797	22	AAU04451
20	253	8.0	797	22	AAU04451
21	253	8.0	797	22	AAU04451
22	253	8.0	797	22	AAU04451
23	253	8.0	797	22	AAU04451
24	253	8.0	797	22	AAU04451
25	253	8.0	797	22	AAU04451
26	253	8.0	797	22	AAU04451
27	253	8.0	797	22	AAU04451
28	253	8.0	797	22	AAU04451
29	253	8.0	797	22	AAU04451
30	253	8.0	797	22	AAU04451
31	253	8.0	797	22	AAU04451
32	253	8.0	797	22	AAU04451
33	253	8.0	797	22	AAU04451
34	253	8.0	797	22	AAU04451
35	253	8.0	797	22	AAU04451
36	253	8.0	797	22	AAU04451
37	253	8.0	797	22	AAU04451
38	253	8.0	797	22	AAU04451
39	253	8.0	797	22	AAU04451
40	253	8.0	797	22	AAU04451
41	253	8.0	797	22	AAU04451
42	253	8.0	797	22	AAU04451
43	253	8.0	797	22	AAU04451
44	253	8.0	797	22	AAU04451
45	253	8.0	797	22	AAU04451

ALIGNMENTS

RESULT 1

AAV99623

ID AAV99623 standard; Protein: 609 AA.

XX AC AAV99623;

XX 03-JAN-2001 (first entry)

XX Neisseria meningitidis BASB040 putative protein sequence #1.

XX BASB040; bacterial disease; respiratory tract infection; bacteraemia;

XX meningitis; cancer; autoimmune disease.

XX Neisseria meningitidis.

XX WO2000034480-A1.

XX 15-JUN-2000.

XX 02-DEC-1999; 99MO-EP095560

XX 07-DEC-1998; 98GB-0026886.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

XX WPI; 2000-423426/36.

XX N-PSDB; AAA48507.

XX Novel BASB040 polypeptides of Neisseria meningitidis useful for

XX diagnostic, prophylactic and therapeutic purposes against microbial

XX diseases comprise a specific amino acid sequence

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

12/2/97
12/7/98

```
XX PS Claim 4; Page 59-60; 98pp; English.
XX OS
XX CC The present sequence is a putative version of the Neisseria
CC meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
CC similar to the D15 outer membrane protein of the bacterium. The protein,
CC its gene, antibodies, antagonists and agonists can be used to diagnose
CC and treat bacterial diseases such as those leading to upper respiratory
CC tract infections, bacteraemia and meningitis. In addition, they can be
CC used in vaccines for use against cancer and autoimmune diseases.
XX SQ Sequence 609 AA;
Query Match 100.0%; Score 3173; DB 21; Length 609;
Best Local Similarity 100.0%; Pred. No. 8.2e-252;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLIKPTALLLPALFFPHAYAPADLSENKAAGFALPKNKSPDTESVKLPKFPVRIDTQ 60
Db 1 MLIKPTALLLPALFFPHAYAPADLSENKAAGFALPKNKSPDTESVKLPKFPVRIDTQ 60
Qy 61 DSEIKDMVEHLPLITQOOEEVLDEKQTGFLAEAPDNVKTMLRSKGYFSSKVSUTEKDG 120
Db 61 DSEIKDMVEHLPLITQOOEEVLDEKQTGFLAEAPDNVKTMLRSKGYFSSKVSUTEKDG 120
Qy 121 AYTVHTTGPRTKIANVGVAJLGDILSDGNLAEYRNALENWQOPVGSDFDQDWSHNSKT 180
Db 121 AYTVHTTGPRTKIANVGVAJLGDILSDGNLAEYRNALENWQOPVGSDFDQDWSHNSKT 180
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Qy 241 SGLARFQPGTPTDLDLDFQALEQNGHYSGASVQADFRLQGDVPPKVSVEVKRHK 300
Db 241 SGLARFQPGTPTDLDLDFQALEQNGHYSGASVQADFRLQGDVPPKVSVEVKRHK 300
Qy 301 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 360
Db 301 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTQNLKRAFGSGGIWYVRORAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTQNLKRAFGSGGIWYVRORAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKROLNNLHPNGHYLDGKIGTTLGTFLSSTALIRT SARAGYFTTPENKKLGTF 480
Db 421 LTASWKROLNNLHPNGHYLDGKIGTTLGTFLSSTALIRT SARAGYFTTPENKKLGTF 480
Qy 481 IIRGQAGYTVARNADVP SGLMFRSGGASSVRGYELDSIGLAPNGSVLPERALLVGSLE 540
Db 481 IIRGQAGYTVARNADVP SGLMFRSGGASSVRGYELDSIGLAPNGSVLPERALLVGSLE 540
Qy 541 YOLPFTRTLSGAVFHDMDGAAANFKRMKLKHGSLGVYRWFSLAPFSFDIAYCHSDKKIR 600
Db 541 YOLPFTRTLSGAVFHDMDGAAANFKRMKLKHGSLGVYRWFSLAPFSFDIAYCHSDKKIR 600
Qy 601 WHISLCTRF 609
Db 601 WHISLCTRF 609
RESULT 2
AAY99624
ID AAY99624 standard; Protein; 609 AA.
XX AC
XX AC AAY99624;
XX DT 03-JAN-2001 (first entry)
XX DE Neisseria meningitidis BASB040 putative protein sequence #2.
XX KW BASB040; bacterial disease; respiratory tract infection; bacteraemia;
```

```
KW meningitis; cancer; autoimmune disease.
XX OS Neisseria meningitidis.
XX PN WO200034480-A1.
XX PD 15-JUN-2000.
XX PF 02-DEC-1999; 99WO-EP09550.
XX PR 07-DEC-1998; 98GB-0026886.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J;
XX DR WPI: 2000-423426/36.
XX DR N-PSDB; AAA48508.
XX PT Novel BASB040 polypeptides of Neisseria meningitidis useful for
XX PT diagnostic, prophylactic and therapeutic purposes against microbial
XX PT diseases comprise a specific amino acid sequence.
XX PS Claim 4; Page 60-61; 98pp; English.
XX CC The present sequence is a putative version of the Neisseria
XX CC meningitidis strain ATCC13090 BASB040 protein sequence. This protein is
XX CC similar to the D15 outer membrane protein of the bacterium. The protein,
XX CC its gene, antibodies, antagonists and agonists can be used to diagnose
XX CC and treat bacterial diseases such as those leading to upper respiratory
XX CC tract infections, bacteraemia and meningitis. In addition, they can be
XX CC used in vaccines for use against cancer and autoimmune diseases.
XX SQ Sequence 609 AA;
Query Match 100.0%; Score 3173; DB 21; Length 609;
Best Local Similarity 100.0%; Pred. No. 8.2e-252;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLIKPTALLLPALFFPHAYAPADLSENKAAGFALPKNKSPDTESVKLPKFPVRIDTQ 60
Db 1 MLIKPTALLLPALFFPHAYAPADLSENKAAGFALPKNKSPDTESVKLPKFPVRIDTQ 60
Qy 61 DSEIKDMVEHLPLITQOOEEVLDEKQTGFLAEAPDNVKTMLRSKGYFSSKVSUTEKDG 120
Db 61 DSEIKDMVEHLPLITQOOEEVLDEKQTGFLAEAPDNVKTMLRSKGYFSSKVSUTEKDG 120
Qy 121 AYTVHTTGPRTKIANVGVAJLGDILSDGNLAEYRNALENWQOPVGSDFDQDWSHNSKT 180
Db 121 AYTVHTTGPRTKIANVGVAJLGDILSDGNLAEYRNALENWQOPVGSDFDQDWSHNSKT 180
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Qy 241 SGLARFQPGTPTDLDLDFQALEQNGHYSGASVQADFRLQGDVPPKVSVEVKRHK 300
Db 241 SGLARFQPGTPTDLDLDFQALEQNGHYSGASVQADFRLQGDVPPKVSVEVKRHK 300
Qy 301 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 360
Db 301 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 360
Qy 361 SNVSYNRSTQNLKRAFGSGGIWYVRORAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Db 361 SNVSYNRSTQNLKRAFGSGGIWYVRORAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Qy 421 LTASWKROLNNLHPNGHYLDGKIGTTLGTFLSSTALIRT SARAGYFTTPENKKLGTF 480
Db 421 LTASWKROLNNLHPNGHYLDGKIGTTLGTFLSSTALIRT SARAGYFTTPENKKLGTF 480
Qy 481 IIRGQAGYTVARNADVP SGLMFRSGGASSVRGYELDSIGLAPNGSVLPERALLVGSLE 540
Db 481 IIRGQAGYTVARNADVP SGLMFRSGGASSVRGYELDSIGLAPNGSVLPERALLVGSLE 540
```

Db 481 IIRGOAGYTVARDNADVPISGLMFRSGASSVRCYELDSICLAGPNCISVLPERALLVGSLE 540
Qy 541 YQLPFTRTLISGAVFHDMDGAAANFKRMKLGSGLVGRWFSPLAPSFDIAYGHSDKKIR 600
Db 541 YQLPFTRTLISGAVFHDMDGAAANFKRMKLGSGLVGRWFSPLAPSFDIAYGHSDKKIR 600
Qy 601 WHISLGTRF 609
Db 601 WHISLGTRF 609

RESULT 3
AA74862
ID AA74862 standard; Protein; 615 AA.

AC AA74862;
DT 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1198.
DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.

OS Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 09-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 98US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;

XX WPI: 2000-062150/05.

XX N-PSDB; AA253624.

XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 669; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.

XX Sequence 615 AA;

Query Match 99.48; Score 3153; DB 21; Length 615;
Best Local Similarity 99.38; Pred. No. 3.7e-250;
Matches 605; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MMIKPTALLPALPEFFHAYAPAAADLSNKAAGFALFKKNSPDTSVKLKPFPVRIDTQ 60
Db 7 MM1KPTALLPALPEFFHAYAPAAADLSNKAAGFALFKKNSPDTSVKLKPFPVRIDTQ 66
Qy 61 DSEIKDMVEEHLPLITQQOEVEVDKEQTGFLEAEAPDNVKTMLRSKGYESSKVSLETKDG 120
Db 67 DSEIKDMVEEHLPLITQQOEVEVDKEQTGFLEAEAPDNVKTMLRSKGYESSKVSLETKDG 126
Qy 121 AVTVHITPGPRTKIANVGVAIILGDILSDGNLAEEYRNALENWQOPVGSFDDQDSWENSKT 180
Db 127 AVTVHITPGPRTKIANVGVAIILGDILSDGNLAEEYRNALENWQOPVGSFDDQDSWENSKT 186
Qy 181 SVLGAVTRKGYPLAKLGNTRAANVPDPTATVLDNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 187 SVLGAVTRKGYPLAKLGNTRAANVPDPTATVLDNVVDSGRPIAFGDFEITGTQRYPEQIV 246
Qy 241 SGLARFQPGTPTDLDLLDFQOALEONGHYSGASVQADFRLQGDVPPVKVSVTEYKRHK 300
Db 247 SGLARFQPGTPTDLDLLDFQOALEONGHYSGASVQADFRLQGDVPPVKVSVTEYKRHK 306
Qy 301 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 360
Db 307 LETGIRLDSYGLGKIAIDYNNLFNKGYSVGVVWMDKYETTTLAAGISQPRNRYGNWT 366
Qy 361 SNVSNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 426
Qy 421 LTASWKROLLNNVLPENGHYLDGKIGTGTTLSTALIRT SARAGYFTTPENKKGITG 480
Db 427 LTASWKROLLNNVLPENGHYLDGKIGTGTTLSTALIRT SARAGYFTTPENKKGITG 486
Qy 481 IIRGOAGYTVARDNADVPISGLMFRSGASSVRCYELDSICLAGPNCISVLPERALLVGSLE 540
Db 487 IIRGOAGYTVARDNADVPISGLMFRSGASSVRCYELDSICLAGPNCISVLPERALLVGSLE 546
Qy 541 YQLPFTRTLISGAVFHDMDGAAANFKRMKLGSGLVGRWFSPLAPSFDIAYGHSDKKIR 600
Db 547 YQLPFTRTLISGAVFHDMDGAAANFKRMKLGSGLVGRWFSPLAPSFDIAYGHSDKKIR 606
Qy 601 WHISLGTRF 609
Db 607 WHISLGTRF 615

RESULT 4

AA74861

ID AA74861 standard; Protein; 615 AA.

XX AC AA74861;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria meningitidis ORF 286 protein sequence SEQ ID NO:1196.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.

XX OS Neisseria meningitidis.

XX PN WO9957280-A2.

XX PD 11-NOV-1999.

XX XX 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998: 98US-0098994.
 PR 02-SEP-1998: 98US-0098062.
 PR 09-OCT-1998: 98US-0103749.
 PR 09-OCT-1998: 98US-0103794.
 PR 09-OCT-1998: 98US-0103796.
 PR 25-FEB-1999: 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-PSDB; AA253623.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 2: Page 667; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254743 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 XX Sequence 615 AA:

Query Match 98.8%; Score 3136; DB 21; Length 615;
 Best Local Similarity 98.7%; Pred. No. 9.1e-249;
 Matches 601; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MMTKPTALLLPALFFPPHAYAPADLSENKAAGFALFNKNSPDTSVKLKPFPVRIDTQ 60
 Db 7 MMTKPTALLLPALFFPPHAYAPADLSENKAAGFALFNKNSPDTSVKLKPFPVRIDTQ 66
 Qy 61 DSEIKOMVEEHLPLITQQQEEVLDKEQTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG 120
 Db 67 DSEIKOMVEEHLPLITQQQEEVLDKEQTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG 126
 Qy 121 AYTWHITPGPKTKIANVGVAILGDLSDGNLAERYRNALENWQOPVGSDFDQDSWENSKT 180
 Db 127 AYTWHITPGPKTKIANVGVAILGDLSDGNLAERYRNALENWQOPVGSDFDQDSWENSKT 186
 Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNDPTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
 Db 187 SVLGAVTRKAYPLAKLGNTRAAVNDPTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 246
 Qy 241 SGLARFQPGTPYDLDLLDFOALQONGHYSGASVOADFRLQGRVPRKVSVEVKRHK 300
 Db 247 SGLARFQPCMPYDLDLLDFOALQONGHYSGASVOADFRLQGRVPRKVSVEVKRHK 306
 Qy 301 LETGIRLDSEYGLGKIAIDYNNFNKGYIGSVVMDKYEITTLAAGISOPRNYRNYWT 360
 Db 307 LETGIRLDSEYGLGKIAIDYNNFNKGYIGSVVMDKYEITTLAAGISOPRNYRNYWT 366
 Qy 361 SNYSYRSTQNLKRAFNKAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
 Db 367 SNYSYRSTQNLKRAFNKAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSAVDLGNSHATM 426
 Qy 421 LTASWROLNNVLPENGHYLDGKIGTTLGTFLSSLTALIRTSARAGYFTTPENKLGTF 480
 Db 427 LTASWROLNNVLPENGHYLDGKIGTTLGTFLSSLTALIRTSARAGYFTTPENKLGTF 486

Qy 481 IIRQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLE 540
 Db 487 IIRQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLE 546
 Qy 541 YQLPFTRTLSGAVFHDMDGAAANFKRMKLGKHSGLGVWRWFSPLAPFSDIAYGHSDKKIR 600
 Db 547 YQLPFTRTLSGAVFHDMDGAAANFKRMKLGKHSGLGVWRWFSPLAPFSDIAYGHSDKKIR 606
 Qy 601 WHISLGTGRF 609
 Db 607 WHISLGTGRF 615
 XX
 XX 03-JAN-2001 (first entry)
 XX
 XX Neisseria meningitidis BASB040 putative protein sequence.
 XX
 XX BASB040: bacterial disease; respiratory tract infection; bacteraemia;
 XX meningitis; cancer; autoimmune disease.
 XX
 XX Neisseria meningitidis.
 XX
 XX WO200034480-A1.
 XX
 XX 15-JUN-2000.
 XX
 XX 02-DEC-1999; 99WO-EP09560.
 XX
 XX 07-DEC-1998; 98GB-0026886.
 XX
 XX (SMIK) SMITHLINE BEECHAM BIOLOGICALS.
 XX
 XX Ruelle J;
 XX
 XX WPI: 2000-423426/36.
 XX
 XX N-PSDB; AA48509.
 XX
 XX Novel BASB040 polypeptides of Neisseria meningitidis useful for
 XX diagnostic, prophylactic and therapeutic purposes against microbial
 XX diseases comprise a specific amino acid sequence
 XX
 XX Claim 4; Page 62; 98pp; English.
 XX
 XX The present sequence is a putative version of the Neisseria
 XX meningitidis strain H44/76 BASB040 protein sequence. This protein is
 XX similar to the D15 outer membrane protein of the bacterium. The protein,
 XX its gene, antibodies, antagonists and agonists can be used to diagnose
 XX and treat bacterial diseases such as those leading to upper respiratory
 XX tract infections, bacteraemia and meningitis. In addition, they can be
 XX used in vaccines for use against cancer and autoimmune diseases.
 XX
 XX Sequence 587 AA:
 XX
 XX Query Match 95.1%; Score 3019; DB 21; Length 587;
 XX Best Local Similarity 98.6%; Pred. No. 3.5e-239;
 XX Matches 579; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Qy 23 AADLSENKAAGFALFNKNSPDTSVKLKPFPVRIDTQDSEIKOMVEEHLPLITQQQEEV 82
 Db 1 AADLSENKAAGFALFNKNSPDTSVKLKPFPVRIDTQDSEIKOMVEEHLPLITQQQEEV 60
 Qy 83 LDKEQGTFLAEAPDNVKTMLRSKGYFSKVSLETKDGAYTVHITPGPKTKIANVGVAIL 142
 Db 61 LDKEQGTFLAEAPDNVKTMLRSKGYFSKVSLETKDGAYTVHITPGPKTKIANVGVAIL 120
 Qy 143 GDILSDGNLAERYRNALENWQOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAA 202

Db 121 GDILSDGNLAEEYNNALNNQVPGSDFDQDSWENSKTSVLGAVTRKAYPLAKLGNTOAA 180
QY 203 VNPOTATVDLNNVVDGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLDDFQQ 262
Db 181 VNPOTATADLNNVVDGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLDDFQQ 240
QY 263 ALEQNGHYSGASVQADFDRLQGRVPPKVSVEVKKHKLKLETGIRLDSYGLGGKIAYDYY 322
Db 241 ALEQNGHYSGASVQADFDRLQGRVPPKVSVEVKKHKLKLETGIRLDSYGLGGKIAYDYY 300
QY 323 NLFNKGYSVVDMDKYETTLAAGISQPRNRYGNVWTSNYSNRTSTONLEKRAFSGGI 382
Db 301 NLFNKGYSVVDMDKYETTLAAGISQPRNRYGNVWTSNYSNRTSTONLEKRAFSGGV 360
QY 383 WYVRDRAGIDARLGAELAGRKIPGSDIDLGNSHATMLTASWKRQLNNVLPENGHYL 442
Db 361 WYVRDRAGIDARLGAELAGRKIPGSDIDLGNSHATMLTASWKRQLNNVLPENGHYL 420
QY 443 DKGITGTLGTLSTALIRTSARAGYFTFPENKKLGTFFIRGOAGYTVARNADVPISGLM 502
Db 421 DKGITGTLGTLSTALIRTSARAGYFTFPENKKLGTFFIRGOAGYTVARNADVPISGLM 480
QY 503 FRSGGASSVRGYELDSGLAGPNSVLPERALLVGSLEYOLPFTTLTSGAVFHDMDGAAA 562
Db 481 FRSGGASSVRGYELDSGLAGPNSVLPERALLVGSLEYOLPFTTLTSGAVFHDMDGAAA 540
QY 563 NFKRMKLKHGSGLVRFSPFLAPSFSDIAYGHSDDKKIRWHISLGTRF 609
Db 541 NFKRMKLKHGSGLVRFSPFLAPSFSDIAYGHSDDKKIRWHISLGTRF 587
RESULT 6
AAV74860
ID AAV74860 standard; Protein; 522 AA.
XX AC AAV74860;
XX 21-MAR-2000 (first entry)
XX DE Neisseria gonorrhoeae ORF 286 protein sequence SEQ ID NO:1194.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN W09957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX DR WPI; 2000-062150/05.
XX DR N-PSDB; AA253622.

PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
PS Claim 2; Page 666-667; 1453pp; English.
XX AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides, AA254537 to AA254576 and AA254616 to AA25473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
SQ Sequence 522 AA;
Query Match 45.5%; Score 1443; DB 21; Length 522;
Best Local Similarity 91.7%; Pred. No. 9e-110;
Matches 287; Conservative 2; Mismatches 16; Indels 8; Gaps 2;
QY 1 MMKPTALLLPALFFPHAYAPADLSENKAAGFALFKNKSPPDTSVKLKPKFPYRIDTQ 60
Db 7 MMKPTALLLPALFFPHAYAPADLSENKAAGFALFKNKSPPDTSVKLKPKFPYRIDTQ 66
QY 61 DSEIKDMVEEHLPLITOOOEVLDDKEQTGFLEAEAPDNVKTMLRSKGYPSSKVSLETGDK 120
Db 67 DSEIKDMVEEHLPLITOOOEVLDDKEQTGFLEAEAPDNVKTMLRSKGYPSSKVSLETGDK 126
QY 121 AYTVDHTGPRFTKIANGVAILGDLSDGNLAEEYNNALNNQVPGSDFDQDSWENSKT 180
Db 127 AYTVDHTGPRFTKIANGVAILGDLSDGNLAEEYNNALNNQVPGSDFDQDSWENSKT 186
QY 181 SVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNNVVDGRPIAFGDFEITGTQRYPEQIV 240
Db 187 SVLGAVTRKGYPLAKGNTRAAVNPDTATADLNNVVDGRPIAFGDFEITGTQRYPEQIV 246
QY 241 SGLARFQPGTPYDLDLDDFQQALEQNGHYSGASVQADFDRL-----OGDRVPVKVSV 293
Db 247 SGLARFQPGTPYDLDLDDFQQALEQNGHYSGASVQADFDRLPRGPRPRQSGRRNGQTQF 306
QY 294 TEVKKRHKLETGIR 306
Db 307 TR-NRHPPRFGR 318
RESULT 7
AAV95820
ID AAV95820 standard; Protein; 578 AA.
XX AC AAV95820;
XX 07-NOV-2000 (first entry)
XX DE Haemophilus influenza strain Rd KW20 BASB067 protein.
XX KW BASB067; outer membrane protein; antigen; vaccine; antibiotic;
XX KW antibacterial; screening; infection; diagnosis; therapy.
XX OS Haemophilus influenzae.
XX PH Key Location/Qualifiers
XX FT Peptide 1..22 /label= Signal_peptide
XX FT Protein 23..578 /label= Mature_protein
XX FT Domain 23..236 /note= "N-terminal domain"
XX FT Domain 237..578

Db 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNNKAPKRYADFIKQ 554

QY 403 GRKIPGSDIDLGNHATML--TASWKRQLNNVLPHPENGHYLDGKIGTTLGTFLSSTAL 459

Db 555 YKTDGAD----GSPKGLLYKGTWCGRNKTDSALWPTRG-YLTG-----VNABI 599

QY 460 IRTSARAGYFFTPENK-----KLGTPIRGOAGYTVARDNA-DVPSGLMFRSGGASV 511

Db 600 ALPGSKLQYYSATHNQTFWFLPSLTKFTLMLGGEVGIAGGYGRTKEIPFFENYGGGLSV 659

QY 512 RGYELDSIGLAGPN-----GSVLP-----ERALLVGSLEYOLP---FTRLTSGAVFHDMD 559

Db 660 RGYE---SCTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLFADAGS 716

QY 560 -----AAANFKRM-----KLKHSGLGLGYRWFSPILAFSEFDIAY---G 593

Db 717 VNDGRYPTAAENGNNKSVYSENAHKSTFTNELRYSAAGAVTWLSPLGPKMFYAYPLKKK 776

QY 594 HSDKKIRWHISIGTRF 609

Db 777 PEDEIORFQFQLGTTTF 792

RESULT 10

AAU03958

ID AAU03958 standard; Protein; 792 AA.

XX AC AAU03958;

DT 23-OCT-2001 (first entry)

XX Neisseria meningitidis serogroup A antigenic protein #1.

XX Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;

KW Bacterial infection; baculovirus; yeast.

XX OS Neisseria meningitidis.

XX FH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..792

FT /note= "Mature N. meningitidis serogroup A antigen"

XX W0200138350-A2.

XX 31-MAY-2001.

XX 28-NOV-2000; 2000WO-IB01851.

XX 29-NOV-1999; 99GB-0028197.

PR 09-MAR-2000; 2000GB-0005698.

XX (CHIR-) CHIRON SPA.

PA (STAT-) STATENS INST FOLKEHELSE.

XX Giuliani MM, Pizzo M, Rappuoli R, Holst J;

XX WPI: 2001-381289/40.

DR N-PSDB; AAS07278.

XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria

PT gonorrhoeae, useful in the manufacture of a medicament for treating and

PT preventing Neisserial bacteria infection.

XX Claim 1; Page 66-68; 92pp; English.

XX The sequence represents a Neisseria meningitidis serogroup A 85 kDa

CC antigenic protein. Neisseria meningitidis colonises the pharynx, causing

CC meningitis and, occasionally, septicaemia in the absence of meningitis.

CC This antigenic protein is useful in the manufacture of a medicament for

CC treating or preventing infection due to Neisseria bacteria, such as

CC meningitis and septicaemia. It is also useful as a diagnostic reagent for

CC detecting the presence of Neisseria bacteria or antibodies raised against

CC Neisseria, and as a reagent for raising the antibodies. The Neisserial

CC nucleotide sequences can be expressed in a variety of different

CC expression systems, for example, mammalian cells, baculoviruses, plants,

CC bacteria and yeast.

CC Note: there are two versions of this sequence displayed in the

CC specification (see AAU04451).

XX

SQ Sequence 792 AA;

Query Match 8.1%; Score 256.5; DB 22; Length 792;

Best Local Similarity 22.4%; Pred. No. 5.1e-12;

Matches 138; Conservative 86; Mismatches 243; Indels 149; Gaps 28;

QY 97 DNKWTMLRSGKGYF-----SSKVSULTKDGAVTVHIT--PGPRTKIANGVAILGDLSDG 149

Db 223 EKVTDFYQNGGFDFRILDTDIQNEKTRQTKITVHIEGGRFRWKSVI-----RG 274

QY 150 NLAEYYRNALE-----NMQQPVGSDFOODSWENSKTSVLGAVTRK----GYPLAKLG 197

Db 275 DTNEVPKAELEKLLTMKPKWYE-----ROOMTAVLGEIONRMGSAGYASEI- 322

QY 198 NTRAAVNPDTATVDLNVVVDGSRPIAFGDFEITGTORYPEQIVSGLARFQPGTPYDLDLL 257

Db 323 SVQPLPNACTKTVDVFLHIEPGRKIYVNEIHTONNKTREVVRRRELQWESAPYDTSKL 382

QY 258 LDFQQALEONGHYSGASVOADFRLQG--DRVPVKVSVTEVKRHKLETGIRLDSYGLGG 315

Db 383 QRSKERVLLGYFD--NVQFDVPLAGTPDKVDLNMSLTERSTGSLDLSAGWVQDTGLVM 440

QY 316 KIADYVNLFNKGIYGSVWMDKYETTLLAAGISQPRNRYGNWTS---NVSYN----- 366

Db 441 SAGVSODNLFGTG--KSAALRASRSKTTLNGSL-----FTDPTFTADGSLGYDIYGKAF 494

QY 367 ----RSTTONLEKRAFSGG-----IWVVRDRAGIDA-----RGAEFLAE 402

Db 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNNKAPKRYADFIK 554

QY 403 GRKIPGSDIDLGNHATML--TASWKRQLNNVLPHPENGHYLDGKIGTTLGTFLSSTAL 459

Db 555 YKTDGAD----GSPKGLLYKGTWCGRNKTDSASWPTRG-YLTG-----VNABI 599

QY 460 IRTSARAGYFFTPENK-----KLGTPIRGOAGYTVARDNA-DVPSGLMFRSGGASV 511

Db 600 ALPGSKLQYYSATHNQTFWFLPSLTKFTLMLGGEVGIAGGYGRTKEIPFFENYGGGLSV 659

QY 512 RGYELDSIGLAGPN-----GSVLP-----ERALLVGSLEYOLP---FTRLTSGAVFHDMD 559

Db 660 RGYE---SCTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLFADAGS 716

QY 560 -----AAANFKRM-----KLKHSGLGLGYRWFSPILAFSEFDIAY---G 593

Db 717 VNDGRYPTAAENGNNKSVYSENAHKSTFTNELRYSAAGAVTWLSPLGPKMFYAYPLKKK 776

QY 594 HSDKKIRWHISIGTRF 609

Db 777 PEDEIORFQFQLGTTTF 792

RESULT 11

AA884745

ID AAB84745 standard; Protein; 792 AA.

XX AC AAB84745;

DT 17-SEP-2001 (first entry)

XX Amino acid sequence of a Neisseria gonorrhoeae protein.

DE Serogroup B protein; outer membrane protein; Neisserial infection;

KW vaccine.

XX Neisseria gonorrhoeae.

OS

XX FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "signal peptide"
 FT Protein 22..792
 FT /note= "mature protein"
 XX WO200152885-A1.
 XX 26-JUL-2001.
 XX 17-JAN-2001; 2001WO-IB00166.
 XX 17-JAN-2000; 2000GB-0001067.
 XX 09-MAR-2000; 2000GB-0005699.
 XX (CHIR-) CHIRON SPA.
 XX Pizza M, Rappuoli R, Giuliani M;
 XX WPI; 2001-451895/48.
 XX N-PSDB; AAH42129.
 XX Composition for treating or preventing infection to, detecting, or for
 PT raising antibodies against Neisserial bacteria, comprises an N.
 PT meningitidis serogroup B outer membrane preparation and an immunogenic
 PT component -
 XX Disclosure: Page 65-67; 83pp; English.
 XX The present sequence represents a Neisseria gonorrhea protein. The
 CC protein is used to produce the compositions of the invention. The
 CC specification describes a composition, comprising a Neisseria
 CC meningitidis serogroup B outer membrane preparation and an immunogenic
 CC component. The immunogenic component is protein disclosed in WO99/57280,
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
 CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
 CC a medicament for treating or preventing infection due to Neisserial
 CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
 CC bacteria; or of antibodies raised against Neisserial bacteria; and/or
 CC a reagent which can raise antibodies against Neisserial bacteria. It may
 CC also be used as a vaccine.
 XX SQ Sequence 792 AA;
 Query Match 8.1%; Score 256.5; DB 22; Length 792;
 Best Local Similarity 22.4%; Pred. No. 5.1e-12;
 Matches 138; Conservative 86; Mismatches 243; Indels 149; Gaps 28;
 QY 97 DNVKTLRSKGVF-----SSKVSUTEDGAYTVHIT--PGPRTKIANGVAILGSDILSDG 149
 DB 223 EKYDFYQNNQYDFRILDTDIQTNEDTKITITVHEGGRWKGVS-----EG 274
 QY 150 NLAEYVRNALE-----NQOVPVGSDFQDSWENSKTSVLGAVTRK-----GYPLAKLG 197
 DB 275 DTNEVPKAELEKLLTMKPKWYE-----ROOMTAVLGEIQNRMGSGAGYSEI- 322
 QY 198 NTRAAVNPDTATLVNVVDSGRIFATGDFEITGQRYQPIVSGLARFQPGTPYDLDLL 257
 DB 323 SVQPLPNAGTKTVPVLVHIEPGRIYVNEITHITGNKNTRDEVVRRRLQMESAPYDTSKL 382
 QY 258 LDFQALQENGHVSGASVQADFDRLQ--DRVPVKVSVTEVVKRHKLETGRILDSVGLGG 315
 DB 383 QRSKERVLLGYFD--NVQFADVPLAGTPDKVDLNSLTERSTGSLDSLAGWVQDTGLVM 440
 QY 316 KIADYNNLNFKNYIGSVVWDMDKYETTLAAGISOPRNYRGNTWS-----NVSYN----- 366
 DB 441 SAGVSQDNLFQGTG--KSAALRASRSKTLNGLSLS-----FTDPYFTADGVSGLGYDIYKAF 494
 QY 367 ----RSTTONLEKRAFSG-----IWVVRDRAGIDA-----RLGAFFLAE 402
 DB 495 DPRKASTSVKQYKTTTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYKNAPRYADPIRK 554

QY 403 GRKIPGSDIDLGNSHATML---TASWKRQLLNVLHPENGHYLDGKIGTTLCTFLSSTAL 459
 DB 555 YGKTDGAD---GSFKGLLYKGTVMGRNRTDSASMPTRG-YLTG-----VNAEI 599
 QY 460 INTSARAGYVFFTPENK-----KLGTFIIRGQAGYTVARDNA-DVPSGLMFRSGGASV 511
 DB 600 ALPGSKLOYYSATHNOTWFFPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENYVGGGLGSV 659
 QY 512 RGYELDSIGLAGPN-----GSVLP-----ERALLVGSLEYOLP---PTRLGSAVPHDMD 559
 DB 660 RGYE---SCTLGPVKYDEYGEKISYGGNKANYSALLPMPGAKDARTVRLSLFADAGS 716
 QY 560 -----AAANFKRM-----KLKHGSLGVGRMFSPLAPSFQIAY-----G 593
 DB 717 VWDGRTYTAENGNNKSVYSENAHKSTFTNELRYSAGGAVTWLSPLGPKFSAVPLKKK 776
 QY 594 HSDKKIRWHISLGRTRF 609
 DB 777 PEDEIQRFQOLGTTF 792

RESULT 12
 AAB23786
 ID AAB23786 standard; Protein: 792 AA.
 XX AC AAB23786;
 XX DT 12-JAN-2001 (first entry)
 XX DE Neisseria gonorrhea amino acid sequence.
 XX KW CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic;
 KW Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae;
 KW bactericidal; antibacterial; vaccine; immunostimulatory; infection;
 KW immune response.
 XX OS Neisseria gonorrhea.
 PN WO2000050075-A2.
 PD 31-AUG-2000.
 XX PF 09-FEB-2000; 2000WO-IB00176.
 XX PR 26-FEB-1999; 99US-0121792.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Grandi G, Rappuoli R, Giuliani MM, Pizza M;
 XX WPI; 2001-015529/02.
 PT Immunogenic composition useful for stimulating an immune response in a
 PT mammal against Neisseria infection, comprises Neisseria antigen and an
 PT adjuvant composition comprising an oligonucleotide with a CG motif -
 XX Claim 22; Page 32; 39pp; English.
 The present invention describes an immunogenic composition (I) comprising a Neisseria antigen and an adjuvant composition comprising an oligonucleotide comprising at least 1 CG motif. Also described is an adjuvant composition (II) comprising an oligonucleotide which comprises at least 1 CG motif and a complete Freund's adjuvant (CFA), where the oligonucleotide preferably comprises at least one phosphorothioate bond. AAA92359 to AAA92385 represent specifically claimed oligonucleotides of the present invention. (I) is useful for stimulating an immune response in a mammal, preferably a human, against Neisseria infection, preferably Neisseria meningitidis infection and in the manufacture of a medicament for inducing a protective immune response in a mammal. The present sequence represents the claimed Neisseria gonorrhea amino acid sequence disclosed in GB-9928197.4, which is given in the present invention.

QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP-----FRTTLSGAVFHDMDG----- 559
 Db 664 --SGTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLFADAGSWDGG 721
 QY 560 -----AAANFKRM-----KLKHSGSLGVRWFSPAPSPFSDIAY-----G 593
 Db 722 TYDONSSTATGRRVNIYGAGNTHKSTFTNELRYSGAGAVTWLSPLGPMKFSYAYPLKKK 781
 QY 594 HSDKKIRWHISLGTRF 609
 Db 782 PEDEIQRFQFOLGTTF 797

RESULT 14
 AAU03959
 ID AAU03959 standard; Protein; 797 AA.
 AC AAU03959;
 XX
 DT 23-OCT-2001 (first entry)
 DE Neisseria gonorrhoeae antigenic protein.
 KW Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
 KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.
 XX
 OS Neisseria gonorrhoeae.
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /note= "Signal peptide"
 FT 22..797
 FT /note= "Mature N. gonorrhoeae antigen"
 XX
 PN WO200138350-A2.
 XX
 PD 31-MAY-2001.
 XX
 PF 28-NOV-2000; 2000WO-IB01851.
 XX
 PR 29-NOV-1999; 99GB-0028197.
 XX
 PR 09-MAR-2000; 2000GB-0005698.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (STAT-) STATENS INST FOLKEHELSE.
 XX
 PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
 XX
 DR WPI: 2001-381289/40.
 DR N-PSDB: AAS07279.
 XX
 PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
 PT gonorrhoeae, useful in the manufacture of a medicament for treating and
 PT preventing Neisserial bacteria infection -
 XX
 PS Claim 1; Page 37-39; 92pp; English.
 XX
 CC The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.
 CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the
 CC pharynx, causing meningitis and, occasionally, septicaemia in the absence
 CC of meningitis. This antigenic protein is useful in the manufacture of a
 CC medicament for treating or preventing infection due to Neisseria
 CC bacteria, such as meningitis and septicaemia. It is also useful as a
 CC diagnostic reagent for detecting the presence of Neisseria bacteria or
 CC antibodies raised against Neisseria, and as a reagent for raising the
 CC antibodies. The Neisserial nucleotide sequences can be expressed in a
 CC variety of different expression systems, for example, mammalian cells,
 CC baculoviruses, plants, bacteria and yeast.
 XX
 SQ Sequence 797 AA;

Query Match 8.08; Score 253; DB 22; Length 797;

Best Local Similarity 21.9%; Pred. No. 1e-11;
 Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

QY 97 DNVKTMLRSKGYF-----SSKVSITKDKGAYTVHIT--PCPRTKIANVGVAILGDLSDG 149
 Db 223 EKVTDFYQNNGYDFRILDITQTNEDKTKQTIKITVHEGGRFRMKVSI-----EG 274
 QY 150 NLAEYRNALD-----NWQPVGSDFDSDSWENSKTSVLGAVTRK-----GYPLAKLG 197
 Db 275 DTNEVPKAELEKLLTMKPGKYE-----ROOMTALGEIQNRMGSGAGYAYSEI- 322
 QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGIARPGQTPPYDLLL 257
 Db 323 SVQPLPNAETKTVDVLHIEPGRKIYVNEIHTGNKTRDEVVRRRLQMESAPDTSKL 382
 QY 258 LDFQALQONGHYSGASVOADFRLQG--DRVVPKVSVEVRKHKLETGIRLDSYGLGG 315
 Db 383 QRSKERVELLGYFD--NVQFDVPLAGTPKVDLNLMSLTERSTGSLDLSAGHVQDTGLVM 440
 QY 316 KIAYDYNNLFNKGYSVYVMDMKYETTL-----AAGISQPRNRYRGNTW---- 359
 Db 441 SAGVSODNLFGTG--KSAALRASRSKTTLNGSLSFDPYFTADGVSLGVDYVYKAFDPRK 498
 QY 360 -TSNVSYNRSTTQNLKRAFSGGIWYVDRAGTDA-----RLGAELFLAEGRKI 406
 Db 499 ASTSIKQKTTTAGAGIRMSVPYTEYDRVNFGLVAEHLTVNTYNKAPKHYADPIKKYKT 558
 QY 407 PGSDIDLGNSHATML--TASWKROLLNNLVHPENHYLDGKIGTTLGTLSLALIRTS 464
 Db 559 DGTD---GSPKGLYKGTGVGWRNKTDALWPTRG-YLTG-----VNAETALPGS 604
 QY 465 RAGYFPTPEK-----KLGTFTIRGOAGYTVARDNA-DVPSGLMFRSGGASSYRGVEL 516
 Db 605 KIQYYSATHNQWTFPLSKFTLMLGGEVGIAGYGRTHREIFFENFYGGGLSGYRGYE- 663
 QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP-----FRTTLSGAVFHDMDG----- 559
 Db 664 --SGTLGPKVYDEYGEKISYGGNKKANVSALLFPMPGAKDARTVRLSLFADAGSWDGG 721
 QY 560 -----AAANFKRM-----KLKHSGSLGVRWFSPAPSPFSDIAY-----G 593
 Db 722 TYDONSSTATGRRVNIYGAGNTHKSTFTNELRYSGAGAVTWLSPLGPMKFSYAYPLKKK 781
 QY 594 HSDKKIRWHISLGTRF 609
 Db 782 PEDEIQRFQFOLGTTF 797

RESULT 15
 AAU04451
 ID AAU04451 standard; Protein; 797 AA.
 XX
 AC AAU04451;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Neisseria meningitidis serogroup A antigenic protein #2.
 XX
 KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
 KW bacterial infection; baculovirus; yeast.
 XX
 OS Neisseria meningitidis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /note= "Signal peptide"
 FT 22..797
 FT /note= "Mature N. meningitidis serogroup A antigen"
 XX
 PN WO200138350-A2.
 XX
 PD 31-MAY-2001.
 XX


```

QY 97 DNVKTLMSKGYF-----SSKVSLTEKDGAYTVHIT--PGPRTKIANVGVAILGDLSDG 149
Db 223 EKVTDFYQNNGFDFRILDTDIQNEDEKTKQTKITVHEGGRFWMCKVSI-----EG 274
QY 150 NLAEYYRNALE-----NMQOPVGSDFDQDQSWENSKTSVLGAVTRK-----GYPLAKIG 197
Db 275 DTNEVPKAELEKLLTMKPGKWE-----ROOMTAVLGEIQNRMGSGAGYSEI- 322
QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLAREPQTPYDLDLL 257
Db 323 SVQPLPNAETKTVDPLVHIEPGRKIYVNEIHTGNKTRDEVVVRRELQRMESAPYDTSKL 382
QY 258 LDFQQALEQNGHYSGASVQADFRLQG--DRVPKVSVTEVKRHKLETGIRLDSEYGLGG 315
Db 383 QRSKERVELLGYFD--NVQFDVAVPLAGTPDKVDLNMSLTERSTGSLDLSAGWVQDTGLVM 440
QY 316 KIAYDYNNLFNKGYIGSVVWMDKYETTL-----AAGISQPRNYRGNYW----- 359
Db 441 SAGVSQDNLFGTG--KSAALRASRSKTTLNGSLSFDPYFTADGVSLGYDVYKAFDPRK 498
QY 360 -TSNVSYNRSTTONLEKRAFSGGIWYVDRAGIDA-----RLGAELFLAEGRKI 406
Db 499 ASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHYADFIKKYGT 558
QY 407 PGSDIDLGNSHATML--TASWKROLLNNVLHPENGHYLDGKIGTTLGTFSLSTALIRISA 464
Db 559 DGTG---GSFGKWLKYGTVGMRNKTDUSALWPTRG-YLTG-----VNAEIALPGS 604
QY 465 RAGYFFTPENK-----KLGTFFIRGQAGYTVARDNA-DVPSGLMFRSGGASSVRYGEL 516
Db 605 KLOYYSATHNQTWFPEPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENFYGGGLSVRGYE- 663
QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTLGAVFHDMDG- 559
Db 664 --SGTLGPKVYDEYGEKISYGNKKNKANVSALLPMPGAKDARTVRLSLFADAGSVDGK 721
QY 560 -----AAANFRM-----KLKHGSLGVYRWFSPILAPFSFDIAY-----G 593
Db 722 TYDMSSSATGGRVQNIYGAGNTHKSTFTNELRYSGAGVATWLSPLGPKMFSYAYPLKKK 781
QY 594 HSDKKIRWHISLGRF 609
Db 782 PEDEIQRFQLGTTF 797

RESULT 17
ID AAB84746
AA B84746 standard; Protein; 797 AA.
XX
AC AAB84746;
XX
DT 17-SEP-2001 (first entry)
XX
DE Amino acid sequence of a Neisseria serogroup A protein.
XX
KW Serogroup A protein; outer membrane protein; Neisserial infection;
XX vaccine.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..797
FT /note= "mature protein"
XX
PN WO200152895-A1.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-IB00166.
XX

```

```

PR 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Pizza M, Rappuoli R, Giuliani M;
XX
DR WPI: 2001-451895/48.
DR N-PSDB; AAH42130.
XX
PT Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against Neisserial bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component -
XX
PS Disclosure: Page 71-74; 83pp; English.
XX
CC The present sequence represents a Neisseria serogroup A protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a Neisseria
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO99/57280,
CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PoA,
CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
CC a medicament for treating or preventing infection due to Neisserial
CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
CC bacteria or of antibodies raised against Neisserial bacteria; and/or
CC a reagent which can raise antibodies against Neisserial bacteria. It may
CC also be used as a vaccine.
XX
SQ Sequence 797 AA;

Query Match 8.08; Score 253; DB 22; Length 797;
Best Local Similarity 21.98; Pred. No. le-11;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

QY 97 DNVKTLMSKGYF-----SSKVSLTEKDGAYTVHIT--PGPRTKIANVGVAILGDLSDG 149
Db 223 EKVTDFYQNNGFDFRILDTDIQNEDEKTKQTKITVHEGGRFWMCKVSI-----EG 274
QY 150 NLAEYYRNALE-----NMQOPVGSDFDQDQSWENSKTSVLGAVTRK-----GYPLAKIG 197
Db 275 DTNEVPKAELEKLLTMKPGKWE-----ROOMTAVLGEIQNRMGSGAGYSEI- 322
QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLAREPQTPYDLDLL 257
Db 323 SVQPLPNAETKTVDPLVHIEPGRKIYVNEIHTGNKTRDEVVVRRELQRMESAPYDTSKL 382
QY 258 LDFQQALEQNGHYSGASVQADFRLQG--DRVPKVSVTEVKRHKLETGIRLDSEYGLGG 315
Db 383 QRSKERVELLGYFD--NVQFDVAVPLAGTPDKVDLNMSLTERSTGSLDLSAGWVQDTGLVM 440
QY 316 KIAYDYNNLFNKGYIGSVVWMDKYETTL-----AAGISQPRNYRGNYW----- 359
Db 441 SAGVSQDNLFGTG--KSAALRASRSKTTLNGSLSFDPYFTADGVSLGYDVYKAFDPRK 498
QY 360 -TSNVSYNRSTTONLEKRAFSGGIWYVDRAGIDA-----RLGAELFLAEGRKI 406
Db 499 ASTSIKQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPKHYADFIKKYGT 558
QY 407 PGSDIDLGNSHATML--TASWKROLLNNVLHPENGHYLDGKIGTTLGTFSLSTALIRISA 464
Db 559 DGTG---GSFGKWLKYGTVGMRNKTDUSALWPTRG-YLTG-----VNAEIALPGS 604
QY 465 RAGYFFTPENK-----KLGTFFIRGQAGYTVARDNA-DVPSGLMFRSGGASSVRYGEL 516
Db 605 KLOYYSATHNQTWFPEPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENFYGGGLSVRGYE- 663
QY 517 DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTLGAVFHDMDG- 559
Db 664 --SGTLGPKVYDEYGEKISYGNKKNKANVSALLPMPGAKDARTVRLSLFADAGSVDGK 721

```

Db	275	DTNEVPKAEKLLTMKPGKWE-----ROOMTAVLGEIQNRHMSAGYAYSEI-327
Qy	198	NTRAAVNPDTATVDLVNVVDSGRPIAEGDFEITGTQRYPEQIVSGLARFQGTPTDLDLL257
Db	323	SVQPLPAETKTVDVFLHTIPGRKIYVNEIHITGNKTRDEVVRRELQMESAPYDTSKL382
Qy	258	LDPQOALEONGHYSASVOADFRLQG--DRVPVKVSVTEVKRIHLETGIRLDSYGLG315
Db	383	QRSKERVLLGYFD--NYOFDAVPLACTPKDKVDLNNLSLTERSTGSLDLASAGWODTGLVM440
Qy	316	KIADYDNLFNKGITGYSVWDMDKYETTL-----AAGISQPRNYRGNW---359
Db	441	SAGVSQDNLFGTG--KSAALRASRSKTTLNGLSLSTFDPYFTADGVSLGYDYGKAFDPRK498
Qy	360	-TSNSYNRSTQNLKRAPFSGGTWYVRDAGIDA-----RLGAEPFLAEGRKI406
Db	499	ASTSIKQYKTKTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVNTYKAPKHYADFLKKYGKT558
Qy	407	PGSDIDLGNSHATML--TASWKROLLNNVLRHPENGHYLDGKITGTTGLTFLSSTALIRTS464
Db	559	DGTD--GSFKWLYKGTGVGWRNKTTDSALWPTRG-YLTG-----VNAEIALPGS604
Qy	465	RAGYFTFPENK-----KLCTFIIRCOAGYTVARONA--DVPSGLMFRSGGASVRYEL516
Db	605	KLQYYSATHINQTFWFLPSKFTFLMLGGEVGIAGGYRTNKEIPFPENFYGGGLGSGVRYE-603
Qy	517	DSIGLAGPN-----GSVLP-----ERALLVGSLEYQLP---FTRTLSCAVFHDMDG-559
Db	664	--SCTLGPVKVYDEYGEKISYGGNKKANVSABLLFPPMPGAKDARTVRLSLFADAGSVWDGK721
Qy	560	-----AAANPKRM-----KLKHSGSLGYRWFSPFLAPFFSDIAY---G593
Db	722	TYDNNSSATGGRVNIYAGNTHKSTFTTWELKYSAGGAVTWLSPLGPMKFSYAYPUKKK781
Qy	594	HSKKIRWRHISLGRFR609
Db	782	PEDEIQRFQQLGTTFF797
RESULT_19		
AAB23788		
ID	AAB23788 standard; Protein; 797 AA.	
XX	AAB23788;	
AC	AAB23788;	
XX	12-JAN-2001 (first entry)	
DT	Neisseria meningitidis serogroup A amino acid sequence.	
XX	CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic;	
XX	Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae;	
XX	bactericidal; antibacterial; vaccine; immunostimulatory; infection;	
XX	immune response.	
OS	Neisseria meningitidis.	
XX	W0200050075-A2.	
PN	31-AUG-2000.	
PD	09-FEB-2000; 2000WO-IB00176.	
XX	26-FEB-1999; 99US-0121792.	
XX	(CHIR-) CHIRON SPA.	
PA	Grandi G, Rappuoli R, Giuliani MM, Piazza M;	
PI	WPI; 2001-015529/02.	
XX	Immunogenic composition useful for stimulating an immune response in	
PT	mammal against Neisseria infection, comprises Neisseria antigen and a	
PT	adjuvant composition comprising an oligonucleotide with a CG motif	

```
Query Match      8.0%; Score 253; DB 22; Length 797;
Best Local Similarity 21.9%; Pred. No. le-11;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;
```

QY	97	DNVTMLRSKGYF----	SSKVSLETKDGATVHIT--	PGPRKIINVGVAILGDLTSDG	149
	:	:	: :: :: :	: :: :: :: :: :	:
Db	223	EKVTFVNNGYDFRILDTDQTWEDKTGITIHVEGRFRWGKYSI-	-----EG	274	
	:	:	:	:	:
QY	150	NLAEYYRNALE-----	NNQQPVGSDFDDDSWENSKTSYLGAATFK---	CYPPLAKIG	197
	:	: :	: :	: :: :: :: :	:


```

Db 499 ASSTSIQYTTTAGAGIRMSVPVTEYDRVNFGLVAEHLVTNTYNKAPKHYADFIKKYGT 558
QY 407 PGSDIDLGNHSHATML--TASWKRLQLNVLNHPENGHYLDGKIGTTLGTFLSSTALIRISA 464
Db 559 DGTD--GSGFKGLYKGTVGWRNKRTDSALMPTRG-YLTG-----VNAEIALPGS 604
QY 465 RAGYFFTPENK-----KLGTFTIRGQAGYTVARDNA-DVPSGLMFRSGGASSVRGYEL 516
Db 605 KLOYYSATHNQWFFPLSKTFTLMLGGEVGIAGGYGRTKEIPFFENFYGGGLGSRGYE- 663
QY 517 DSIGLAGPN-----GSLVP-----ERALLVGSLEYQLP---FTRTLGSAGVPHDMGD----- 559
Db 664 --SGTLGPKVYDEYGEKISYCGNKKANVSABELLPMPGAKDARTVRLSLFADAGSVMWCK 721
QY 560 -----AAANFKRM-----KLKHGSLGVRWFSP LAPSFEDTAY-----G 593
Db 722 TYDNDSSSATGGRVONIYGAGNTHKSTFTNELRYVSAGGATVWLSPLGPKFRVAYPLKKK 781
QY 594 HSDKKIRWHISLGRTRF 609
Db 782 PEDEIQRFQQLGTTTF 797

RESULT 21
AAG78603
ID AAG78603 standard; Protein; 896 AA.
XX
AC AAG78603;
XX
DT 20-NOV-2001 (first entry)
XX
DE Lawsonia intracellularis protein SEQ ID NO: 5.
XX
KW HtrA; PonA; HlyC; YefW; ABC1; Omp100; Lawsonia intracellularis infection;
XX vaccine.
XX
OS Lawsonia intracellularis.
XX
PN JP2001169787-A.
XX
PD 26-JUN-2001.
XX
PF 20-OCT-2000; 2000JP-0320736.
XX
PR 22-OCT-1999; 99US-0160922.
XX
PA {PF12 } PFIZER PROD INC.
XX
PS WPI; 2001-592540/67.
XX
PT Lawsonia intracellularis polynucleotide and encoded protein, used to
prevent Lawsonia intracellularis infection -
XX
PS Claim 12; Page 43-45; 67pp; Japanese.
XX
CC The present invention provides isolated polynucleotides encoding HtrA,
PonA, HlyC, Lyss, YefW, ABC1 or Omp100 protein of Lawsonia
intracellularis. The sequences can be used in vaccines for the prevention
of Lawsonia intracellularis infection. The present sequence is a protein
of the invention.
XX
SQ Sequence 896 AA;

Query Match 7.8%; Score 247; DB 22; Length 896;
Best Local Similarity 22.8%; Pred. No. 3.7e-11;
Matches 140; Conservative 95; Mismatches 276; Indels 104; Gaps 25;

QY 64 IKDWVEHLPLI---TQOQEEVLDKEQ-----TGFLAEAPDN---VKTMLRSKGYF 109
Db 317 IKDVRLEGLETKATLKKLEALUERNFLSWFTGTGTVLREYLERDSIAISAYAMHGVK 376
QY 110 -----SSKVSLETKDGAYTVHTPEPRTKIANVGAVTILGSDGILSDGNLAERYRNAL----- 159

```

```

Db 377 DIOVAPEVTFNEKGIIVITFRVKEGKRYKI-----GKIDFKGDLIETNEQLLKVTKI 428
QY 160 ---ENWQQVGSDFDQDQSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVY 216
Db 429 DDHKNYEQYFSLSVNQDDVK---ALTDFYSDYGYAEV--DLETTKNEEDATIDVTFLL 483
QY 217 DSGRPIATFGDFFETGTQRYPEQIVSGIARFQGPYPYDLDLLDFQOALEQNGHYSGASVO 276
Db 484 DKQKQVFLRIIVEGNTTRDNVILRELRLADGDLFNGQHLRNSNECLNRLGYFN---Q 539
QY 277 ADPRL---QGRDVPVKVSVTEVRRHKLETGJ---RLDSEYGLGKGIAYDYNNLFNKGYIG 331
Db 540 VDTDTLPTGKDDDEVLLVKVQEARTGATGGVGYSTHKKFVSGSIS--ERNLMGKGYLL 597
QY 332 SVYWDMDKYETTLLAAGISOPRNYRCNWTNSVSYN-RSTTONLEKRAFSGGIWYVDRAG 390
Db 598 SIETCFISSKSSSLDLSFTNPRVYDTDFGFSNNIYTLDEWDDDFRKKTYG-----DTIR 650
QY 391 IDARLGA-EFLAEGRKI-----PGSDIDL-GNSHATMLTASMKRQLNVLHP 436
Db 651 LFHPIGEYSSIFVGYRIDQYRLYDIPSTAPRSYLDYQCKNITSSVWVGGETFSDTSRERP 710
QY 437 ENGH-----YLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTFLIRGQAGYT 489
Db 711 SKGHIAKLIVEYGGGLG-----GNDNFFKPIAELOQGYISRSKNNIILHWRTRAGAA 763
QY 490 VARDNADVPGLMFRSGGASSVRGYELDSIGLAGPN-GSVLPERALLVGSLEYQLPPTRT 548
Db 764 YKNSKKPVVDFRFFIGGIDSTIRGYDTEDLAPKDPFRFCDEIGDRMAFLNLEY IWTFOPE 823
QY 549 LSGAV--PHDMGDAANFKR-----MKLKHGSLGVRWFSP LAPSFEDTAYG-----HSD 596
Db 824 LGLALVPFDIGCFOTDSVOTSNPFSKLLKQSYGLERLHRSFPGDLRF--AYGIPLAKNVS 881
QY 597 KKI--RWHSISLGRTRF 609
Db 882 KKTGRGRFESMGQFF 896

RESULT 22
ABG17710
ID ABG17710 standard; Protein; 474 AA.
XX
AC ABG17710;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17701.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS81697.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations

```


Db 488 FENYDNKSDTSSNYKRTTYGSSNVTGLFPVNNNSYYVGLGHTYNTKISNFALEYNRLYI 547
QY 403 -----GRKIPSDIDLGNSHATMLTASWKRQLLNYYLHPENGHYLDGKIGTTLGTFLLS 456
Db 548 QSMKFGNGIKTNDFD-----FSGWNYNSLNRGYFFTKG--VKASLGRVITPGSD 597
QY 457 TALIRTSARAGYFFFTPENKLGTFILRQAGYTVARDNADVP SGLMFRSGGASSVRGYEL 516
Db 598 NKYYKLSADVQGYFLDRHLWVWSAKASAGYANGFGNKRLLPFYQYTAGGIGSLRGFAY 657
QY 517 DSTGLAGPN-----GSVLPERALLVGSLEYOLP-----FTRT----- 548
Db 658 GSI---GPNAIYOGNNKFNKISSDVIGGNAIATASAEILVPTPFVSDKSQNTVTSLEF 714
QY 549 -----LSGAVFHDWGDAAAANFRKMLKHGSLGVLGRWFSPLAPFSDIA- 591
Db 715 DAASVWNTKWKSDKNLSDLPDYG---KSRTRASTGVGFQWQSPSPGVVFSYAK 771
QY 592 -----YCHSDKKIRWHISLGTRE 609
Db 772 PIKKYENDDOVE-QFOFSIGGSF 792

RESULT 27
AAR53755
ID AAR53755 standard; Protein; 797 AA.
AC AAR53755;
XX XX
DT 08-DEC-1994 (first entry)
DE XX
DE H. Influenzae b Eagen D15 sequence.
XX XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; Hib.
XX XX
OS Haemophilus Influenzae type b Eagen strain.
PN XX
PN W09412641-A.
XX XX
PD 09-JUN-1994.
XX XX
PE 23-NOV-1993; 93WO-CA00501.
XX XX
PR 23-NOV-1992; 92GB-0024584.
XX XX
PA (CONN-) CONNAUGHT LAB LTD.
XX XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX XX
DR WPI: 1994-200269/24.
DR N-PSDB; AAO66199.
XX XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus Influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX XX
PS Disclosure; Fig. 1B; 161pp; English.
XX XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC Chromosomal libraries of H. Influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAO66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX XX
SQ Sequence 797 AA;

Query Match 6.6%; Score 209.5; DB 15; Length 797;

Best Local Similarity 20.0%; Pred. No. 3.7e-08;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSEIKDMVEEHLPLITQOOEEVLDEKQTFGLAEAPONVATMLRSKGYFSSKYSLT---- 116
Db 184 ESVSSSTLQEOEMLQPDSSWMLGNKPEGAQFEXDLQSIIRDYVLLNNGYAKAQITKTVDQL 243
QY 117 --EKDCA-VTVHTTPGPRTK-----IANVG--VAILGDILSDGNLAERYNALENWQOP 165
Db 244 NDEKTVKVVYTIQVNGLOQLDLSARIIGNLGMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDFQDDSWENSKTSVIGAVTRKGYPLAKIGNTRAAVNP-----TATVDLVNVVDSGRP 221
Db 297 -----DIADVENAIAKILG---ERGYGSATVNSV-----PQFDDANKTLAITLVVDAGR 343
QY 222 IAFDFEITGTQYRPEQIVSGLARFPQCTPYDLDLLDFOALEQNGHTSGASVOADPDR 281
Db 344 LTVQLRFEQNTVSADSTLRQEMRQOEGTWNLSQVELGKIRLDRGTGFFE--TVENRIDP 401
QY 282 LQG--DRVPVKVSVTEVRKHLEGTGIRLDSYGLGGKIADYVYNLFNKGYIGSVWMDMK 339
Db 402 INGSNDVDVYVKERNTGSTINFQIGYGTESGISYQASVKQDNFLGTGAASVIACTKND 461
QY 340 YETTLAAGISQPR-----NYRGNVTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
Db 462 YGTSVNLGYTEPYFTKDGVS LGGVFFPENYDNSKSDTSSNKRTTYGSNVTGLFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKRQLL 430
Db 522 SYVYGLGHTYNTKISNFALEYNRLYIQSMKFGNGIKTNDPD-----FSGWNYNSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARACYFTTPENKLGITFIIRGOAGYTV 490
Db 574 NRGYFTPKG--VKASLGRVITPGSDNKYYKLSADVOGFPFLDRDHLWVWSAKASAGYAN 531
QY 491 ARDNADVPSGLMFRSGGASSVRGYELDSIGLAGPN-----GSVLPPE 531
Db 632 GFGNKRLLPFYQYTAGGIGSLRGFAYGSI---GPNAIYAEYGNSGCTGTGFKKISSDVIGG 688
QY 532 RALLVGSLEYOLP-----FTRTLGSAVFHDM-----GDAANFKRW---- 567
Db 689 NAIATASAEILVPTPFVSDKSQNTVTSLEFVDAASVWNTKWKSDKNLSDVLRPLPDYG 748
QY 568 ---KLKHGSLGVWRVPSPLAPFSDIA-----YGHSDKKIRWHISLGTRE 609
Db 749 KSSRIRASTGVGFQWQSPIGPLVFSYAKPIKKYENDDOVE-QFOFSIGGSF 797

RESULT 28
AAR53756
ID AAR53756 standard; Protein; 797 AA.
XX XX
AC AAR53756;
XX XX
DT 08-DEC-1994 (first entry)
DE XX
DE H. Influenzae b Minn A D15 sequence.
XX XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; NTHi.
XX XX
OS Haemophilus Influenzae type b Minn A strain.
PN XX
PN W09412641-A.
XX XX
PD 09-JUN-1994.
XX XX
PF 23-NOV-1993; 93WO-CA00501.
XX XX
PR 23-NOV-1992; 92GB-0024584.
XX XX
PA (CONN-) CONNAUGHT LAB LTD.

Chong P., Klein M., Loosmore S., Sia DYC, Thomas W;
Yang Y;
WPI; 1994-200269/24.
N-PSDB; AAQ66200.
Nucleic acid encoding D15 outer membrane protein - esp. of
Haemophilus influenzae, and related proteins, vectors, antisera
etc. useful in vaccines, for diagnosis and for passive
immunisation.
Disclosure: Fig. 1C; 161pp: English.
Outer membrane protein (OMP) D15 genes were isolated by screening
chromosomal libraries of H. influenzae type b (Hib) strains Ca,
Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
12055. Nucleotide sequences were determined for the D15 genes
(AAQ66198-202) and the corresponding aa sequences were derived
(AAQ33754-58). D15 OMP can be produced easily and on a large scale,
free of other antigens and lipooligosaccharides, by recombinant DNA
methods using the isolated genes.
Sequence 797 AA;
Query Match
Best Local Similarity 5.68; Score 208.5; DB 15; Length 797;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSEIKDMVEHLPLITOOQEEVLDKEQTGFLEAEAPDNVATMLRSKGYSSKVSILT--- 116
DB 184 ESVSSSTLQEQMELQPDSSWKLGNKFEQAQFEKDLQSI R D Y Y L N N G Y A K A Q I T K D V Q L 243
QY 117 --EKDGA-YTVHITGPRTK-----TANWG--VAILGDI L S D G N L A E Y Y R A L E N W Q O P 165
DB 244 NDEKTKVNTIDVNGELQYDLSRARIIGNLGMSAELEPLLSALHLNDFRRS----- 296
QY 166 VGSDFDQDSDWNSKTSVLGAVTRKGYPLAKLGNTRAAVNDP-----TATVDLNVVVVDSGRP 221
DB 297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDVKTKLTLAVVDAGRR 343
QY 222 IAFGDFEITGTORYPEQIVSGLARFQGTPTVDLIDLLDFQALEONGHYSGASVOADPDR 281
DB 344 LTVRQLRFEGNTVSADSLRQMQRQEGTWTNSQLVELGKIRLDRTGFEE--TVENRIDP 401
QY 282 LQG--DRVPEKVSVEYVRKHLEGTIRLDSYGLGGKIAYDYTNLFNPKGYIGSVYMDMK 339
DB 402 INGSNDVDVWYVKVERNTGSINFGYGTSGISYQASVKQDNFLGTGAASVIACTKND 461
QY 340 YETTLAAGISOPR-----NYRGNWFTSNVSNYNS--TQNLKRRASGGI---WYVDRDA 389
DB 462 YGTSVNLGTEPTFTFKDGVSLGGNVFEENDNSKSDTSNFKRTTYGTSNVTLCGFPVNNNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASMKRQLL 430
DB 522 SYVVLGHTYNTKISNFALEYNRLYIOSMKFKNGIKTNDFD-----FSGWNYNSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFSLSTALIFTSARACYFFTPENKKLGTFIIRGOAGYTV 490
DB 574 NRGYFTPKC--VKASLGGRTVLPGSDNKYKLSADVQGYPLDRDHLWVVSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASVGRGYELDSIGLAGPN-----GSLVPE 531
DB 632 GFGNKKRLPEQYTVTAGGICSLRGFAYGSI--GPNAIYAEGVCGSGTGFKKISSDVIGG 688
QY 532 RALLVGSLEYQLP-----FTRLTSLGAVFHDN-----GDAANFKRM----- 567
DB 689 NATATASAEIIVPTFPVSDKSONTVRTSLPVDAAVSVNTKWKSDKNGLSDVLKRLPDPYG 748
QY 568 ---KCLKHSGSLGVRWFSPLAPSFEDIA-----YGHSDKKIRWHSIGLGRF 609
DB 749 KSSRIIRASTGVGFQMSPTGLPVFSYAKPIKKYENDDVE--OFOSYIGSGSF 797

RESULT 29	
ID	AAR53754
AA	AAR53754 standard; Protein; 797 AA.
XX	
AC	AAR53754;
XX	
DT	08-DEC-1994 (first entry)
XX	
DE	H. influenzae b Ca D15 sequence.
XX	
KW	Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW	OMP; outer membrane protein; Hib.
XX	
OS	Haemophilus influenzae type b Ca strain.
XX	
PN	W09412641-A.
XX	
PD	09-JUN-1994.
XX	
PF	23-NOV-1993; 93WO-CA00501.
XX	
PR	23-NOV-1992; 92CB-0024584.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	
PI	Chong P, Klein M, Loomore S, Sia DYC, Thomas W;
PI	Yang Y;
XX	
DR	WPI: 1994-200269/24.
XX	
DR	N-PSDB; AA066198.
XX	
PT	Nucleic acid encoding D15 outer membrane protein - esp. of
PT	Haemophilus influenzae, and related proteins, vectors, antisera
PT	etc. useful in vaccines, for diagnosis and for passive
XX	immunisation.
XX	
PS	Disclosure; Fig. 1A; 161pp; English.
XX	
CC	Outer membrane protein (OMP) D15 genes were isolated by screening
CC	chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC	Egan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC	12085. Nucleotide sequences were determined for the D15 genes
CC	AA066198-202) and the corresponding aa sequences were derived
CC	(AAR53754- 58). D15 OMP can be produced easily and on a large scale,
CC	free of other antigens and lipooligosaccharides, by recombinant DNA
CC	methods using the isolated genes.
XX	
SO	Sequence 797 AA:
	Query Match 6.5%; Score 207.5; DB 15; Length 797;
	Best Local Similarity 20.0%; Pred. No. 5.5e-08;
	Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps
Qy	61 DSEIKDWEHPLPLTQQOEVLKDEQOTGFLEAPDNVKTMLRSGVFSSKVSLT----
Db	184 ESVSSSTLQEQMELQPSDWKLGKNGKFEQAQFKDLQSIDRYLNNGYAKAQITKTQVQL 24.3%
Qy	117 --EKDGA-TTVHTTGPRTK-----TANVG--VAILGOITLSDGNLAETVYRNALENWQOP 165%
Db	244 NDEKTKVNTIDVNEGLQYDLSARIICNLGGMSAELEPLLSALHLNDFRRS-----296%
Qy	166 VGSDFQDQSWENSKSVLGATVRGKYPLAKIGNTRAIVNP-----TAIVDLNVVVDGSRP 221%
Db	297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343%
Qy	222 IAFGDFEITGTORYPEQIVTSGLARFPQCTPVLDLLDLDFOOALEONGHVS GASVQADFDR 281%
Db	344 LTVROLRFEGNTVTSADSTLRQEMRQECTWYNSQLVELCKIRLDRTGFFE--TVENRDP 401%
Qy	282 LQG--DRVVPKVSVEVTKRHKLETGIRLDSLEYLGKGKIAYDVYVNFNKGYSIGSVVWMDK 339%
Db	402 INGSNDEVDVYKVKERNTGTSINFGIVGYESGTSIGVQASVKQDNFLGTCAAVSITAGTKND 461%

KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
KW cellular immune response.
XX Helicobacter pylori.
OS W09921959-A2.
PN W09921959-A2.
XX PD 06-MAY-1999.
XX PF 28-OCT-1998; 98WO-US22883.
XX PR 17-DEC-1997; 97US-0993001.
XX PR 28-OCT-1997; 97US-0959131.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
XX WPI; 1999-326698/27.
DR N-PSDB: AAX75802.
XX Cellular vaccine against Helicobacter pylori
PS Claim 7; Page 243-247; 352pp; English.
XX The invention relates to a vaccine for preventing or treating infections
CC by Helicobacter pylori. The vaccine contains at least one isolated
CC H. pylori polypeptide, or its fragments, in a carrier, where the
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
CC humoral and cellular immune responses. The vaccines are used to treat or
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
CC (OMPs) AAX17160 to AAX17218.
XX SQ Sequence 925 AA;
Query Match 5.5%; Score 174; DB 20; Length 925;
Best Local Similarity 21.0%; Pred. No. 3.9e-05;
Matches 123; Conservative 72; Mismatches 202; Indels 188; Gaps 23;
QY 186 VTRKGYPLAKGNTRAAVNDTATVNLNVV-----DSGRPIAFGDFEITGTORYPEQI 239
Db 368 IADKGTAF-----AVKPDLDKDKNGLVKVIYRIEYQDMVHNDVIIIGNORTSDRI 420
QY 240 VSGLARFQPTPYDLDLLDQQAQNGHYSGASVQADFRLQGDVRPVKVSVEYKRRH 299
Db 421 IRRELLGPKDKYNLTKLRNSENRLRLGFFS--KVKIEKRVNSSLMDLLVVEEGRTG 478
QY 300 KLETGIRLDSEYGLGGKIADYNNLKNKYIGSVYWD-----MDKYETTLAAG- 347
Db 479 QIQPGIGYSGYGLMLNGVSERNLFTGQSMISLYANIATGGRSYPGMPGAGRMFAGN 538
QY 348 --ISQPR-----NRY-----GNYTSTNVSYNRSTPTQ 371
Db 539 LSLTNPRIFFDSWYSTINLADYRISYQIQGGGFGVNVGRMLGNRTHVSLGYNLNVTK 598
QY 372 NLEKRAFSGGIWYVRDAGID-----ARLGAEEFLAERK--IPGSDIDLG---- 414
Db 599 LL---GFSSPL-YARYYSNVASVPRQCTPASVLIINRLSGRPLPVPESCSPGAITT 654
QY 415 -----NSHATMLTASMKRQL-----LNNVLHPENG-----HYLDGKIGTTLGLFLLST 457
Db 655 SPEIKGIWDRDYHTPTSSFTLDVSYDNTDDYFPRNGVIFSSYATMGLSPSGTLNSWN 714
QY 458 AL---IFTSARAGYFFTPENKKGTFI-----TRGQGYTVARDNAD--VPSGLMFRSGG 507
Db 715 GUGGNRTKVKYGFAYHHQLQKYLIIIDLIARPKTOGGY-IFRYNTDQYLPUNSTFYMCG 773
QY 508 ASSVRGIELDSIGLAPNGSVLPE-----NGSITPKDEPLWLGDDGIFTASTELSYGLKAAKRWLANFFD 531
Db 774 VTTVRGRF-----NGSITPKDEPLWLGDDGIFTASTELSYGLKAAKRWLANFFD 824

OY 532 -----RALLVGSLEYOLPFTRLSCAVPHDMCDAANFKRMKLKHGSLGVRWPSPLAP 585
Db 825 FGFLTFRTPTRGSSFFYNAPT-----ANFKDYGVVGAGFERATWRASTGLQIEWISPMGP 880
OY 586 --FSFDIAY-----GHSDDKKIRWHISLQTRF 609
Db 881 LVLIFPIAFPNOWGDGNGKCKGLCFNPNNDYTOHFEFSMGTRE 925
RESULT 32
ABB90599
ID ABB90599 standard; Protein; 790 AA.
XX AC ABB90599;
XX DT 29-JUL-2002 (first entry)
XX DE Chlamydia pneumoniae cp6576 protein, SEQ ID NO:147.
XX KW Chlamydia pneumoniae cp6576 protein; immunogen; vaccine; diagnosis;
KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029.
XX OS Chlamydia pneumoniae.
XX FH Key Location/Qualifiers
FT Peptide 1..26
FT Protein /label= Signal_peptide
FT 27..790
FT /note= "Mature protein"
XX WO200202606-A2.
XX PD 10-JAN-2002.
XX PF 03-JUL-2001; 2001WO-IB01445.
PR 03-JUL-2000; 2000GB-0016363.
PR 11-JUL-2000; 2000GB-0017047.
PR 21-JUL-2000; 2000GB-0017983.
PR 07-AUG-2000; 2000GB-0019368.
PR 18-AUG-2000; 2000GB-0020440.
PR 14-SEP-2000; 2000GB-0022583.
PR 10-NOV-2000; 2000GB-0027549.
PR 22-DEC-2000; 2000GB-0031706.
XX (CHIR-) CHIRON SPA.
XX PI Ratti G, Grandi G;
XX WPI; 2002-154726/20.
DR N-PSDB; ABL91257.
PT Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes -
XX Claim 1; Page 113; 364pp; English.
XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of

CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX
SQ Sequence 790 AA;
Query Match 5.5%; Score 173.5; DB 23; Length 790;
Best Local Similarity 20.7%; Pred. No. 3.4e-05;
Matches 148; Conservative 92; Mismatches 266; Indels 209; Gaps 33;
QY 22 PAADLSENKAAGFALFKKSPDTSVKLKPKFPVRIQDSEIKDMVEEHLPLITQO--Q 79
Db 87 PKVEFSEGR-TNIALHLIAKPSIRNIHSG-----NQVPEHKILKTLQIYR 132
QY 80 EEVLDEKQTCFLAEAPONVKTMLRSKGYFSSKVSLEKDCGAYTVH-----ITPGPRTK 133
Db 133 NDLFEREK--FL--KGLDRLTYLLKRGYFASSVDYSLEHNOEKGHIDVLKINEGPCGK 188
QY 134 IANVGVAILG-----DILSDGNLA-----EY 155
Db 189 IKOLTSGISRSEKSDIOEFQTKOHSSTTSWFTGAGLYHPDIVEQDSLAITNYLHNNGY 248
QY 156 RNALENWQPVGSDFDQDSWEN-----SKTSVLGAVTRKGYPL----- 193
Db 249 ADAI-----VNSHYDLDKGNILLYMDIDRGSRYTLGHVHIOGFEVLKRLIEKOSQVG 302
QY 194 -----AKLG--NTRAAV-----NPDATVDLVNVDVSGRPTAFGD 226
Db 303 PNDLYCPDKIMWGAHKIKQTVAKYGININTVDVLFIPHATRPIDYTYEVSSEGSPPKVG 362
QY 227 FEITGQRYPEQIVSGLARFQPGTPYDLDLLDFQOALEQNGHYGASV-----QA 277
Db 363 IKITGNTHTKSDVILHETSLRPGDTFENRLKLEDETEQLRNTGYFQSVSVTVRSQDLPWG 422
QY 278 DFDRLQGDVRPVKVSUTE-----VKRHKLETGIRL-DSEYGLGGKATAYDYNLFNKG 328
Db 423 NADQYRDIEVEVKETTGNLGLFGLFSSLDNLFGGIELSESNFDLFGA-----RNIFSKG 477
QY 329 Y-----IGSVVMDMDKYETTLAAGISOPRYNRYNWTSNV-----SYNRST 369
Db 478 FRCLRGGGHFLFKANFGDKVTDYTLKWT-----KP-HFLNTPWILGIELDKSINRAL 529
QY 370 TONLEKRAFSGGI--WYVRD---RAGIDARLGAEFLAEGRK-IPGSDIDLGNSHATMLTA 423
Db 530 SKDYAVOTTGGVNSYTYILNEHLKYGFLYRGQSOTSLHEKRRFLGPNIDSNKGFSAAGV 589
QY 424 SWKRQLNNVLPENG-----HYLDGKIGTTLGTLFSL-STALIRTSARAGYFTFPENKK 476
Db 590 NLMYDSVDSRPTPTTGIRGGVTFFVSGLGTYHFTKLSLNSIYRKLTRK----- 639
QY 477 LGFTIRGQAGTVARDNA---DVPGLMRSSGASVRCGYELDSIG---LAGPNGSVL 529
Db 640 -GLIKTGAQIKPIKYSNTTAEQVPSRPFPLGGETTIVRGKSEFIQPKYSATEPOGGU- 697
QY 530 PERALLVGSLEYOLPFR--TLSCGAFHDMGDAANFKMKLK---HSGSLGVYR 579
Db 698 --SSLLI-SEEFQYPLIRQPMISAFVFLDGSFVGLQBYKISLKLDRSSAGFLGRF 749
RESULT 33
ID AAY34896 standard; Protein: 795 AA.
XX
AC AAY34896;
CC
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.
OS
XX
PN W09927105-A2.
XX
PD 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-1B01890.
XX
XX 04-NOV-1998; 98US-0107078.
PR
PR 21-NOV-1997; 97FR-0014673.
XX
XX (GEST) GENSET.
PA
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 832-834; Disclosure: 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
XX Sequence 795 AA;
SQ
Query Match 5.5%; Score 173.5; DB 20; Length 795;
Best Local Similarity 20.7%; Pred. No. 3.4e-05;
Matches 148; Conservative 92; Mismatches 266; Indels 209; Gaps 33;
QY 22 PAADLSENKAAGFALFKKSPDTSVKLKPKFPVRIQDSEIKDMVEEHLPLITQO--Q 79
Db 92 PKVEFSEGR-TNIALHLIAKPSIRNIHSG-----NQVPEHKILKTLQIYR 137
QY 80 EEVLDEKQTCFLAEAPONVKTMLRSKGYFSSKVSLEKDCGAYTVH-----ITPGPRTK 133
Db 138 NDLFEREK--FL--KGLDRLTYLLKRGYFASSVDYSLEHNOEKGHIDVLKINEGPCGK 193
QY 134 IANVGVAILG-----DILSDGNLA-----EY 155
Db 194 IKOLTSGISRSEKSDIOEFQTKOHSSTTSWFTGAGLYHPDIVEQDSLAITNYLHNNGY 253
QY 156 RNALENWQPVGSDFDQDSWEN-----SKTSVLGAVTRKGYPL----- 193
Db 254 ADAI-----VNSHYDLDKGNILLYMDIDRGSRYTLGHVHIOGFEVLKRLIEKOSQVG 307
QY 194 -----AKLG--NTRAAV-----NPDATVDLVNVDVSGRPTAFGD 226
Db 308 PNDLYCPDKIMWGAHKIKQTVAKYGININTVDVLFIPHATRPIDYTYEVSSEGSPPKVG 367
QY 227 FEITGQRYPEQIVSGLARFQPGTPYDLDLLDFQOALEQNGHYGASV-----QA 277
Db 368 IKITGNTHTKSDVILHETSLRPGDTFENRLKLEDETEQLRNTGYFQSVSVTVRSQDLPWG 427
QY 278 DFDRLQGDVRPVKVSUTE-----VKRHKLETGIRL-DSEYGLGGKATAYDYNLFNKG 328
Db 428 NADQYRDIEVEVKETTGNLGLFGLFSSLDNLFGGIELSESNFDLFGA-----RNIFSKG 482
QY 329 Y-----IGSVVMDMDKYETTLAAGISOPRYNRYNWTSNV-----SYNRST 369
Db 483 FRCLRGGGHFLFKANFGDKVTDYTLKWT-----KP-HFLNTPWILGIELDKSINRAL 534
QY 370 TONLEKRAFSGGI--WYVRD---RAGIDARLGAEFLAEGRK-IPGSDIDLGNSHATMLTA 423

Db 535 SKDYAVQYGVNVTYILNHLKYLFGYRGSTSLHEKRFLLGPNIDSNKGFSAAGV 594
QY 424 SHKROLLNNVLHPENG-----HYLDGKIGTTLGTFELS-STALIRTSARAGYFFTPENKK 476
Db 595 NLNYSVDSPTPTTIGRGGVTFEVSGLGTYHTFKLSLNSISYIKLTKRK----- 644
QY 477 LCTFIIRGQAGYTVARDNA---DVPSGLMFRSGGASSVRGYELDSIG---LAGPNSVL 529
Db 645 -GILKIGEAQFIKPYNTAAGVPVPSERFELGGETTVRGYSFIIGPKYSATEPOGL- 702
QY 530 PERALLVGSLEYQLPFR--TLGAVFHDMDGDAANFKRMKLLK---HSGGLGVRW 579
Db 703 --SSLLI-SEEFYPLIRQPNISAFVLDSEGFVGLQYKISLKLRLSSAGFLRF 754
RESULT 34
ID AAW71477 standard; Protein; 847 AA.
AC AAW71477;
XX
DT 09-NOV-1998 (first entry)
DE Helicobacter polypeptide GHPO 107.
XX
KW GHPO 107; infection; therapy; diagnosis; vaccine; gastritis;
KW ulcer.
XX
OS Helicobacter pylori.
XX
PN W09821225-A1.
XX
PD 22-MAY-1998.
XX
PF 14-NOV-1997; 97MO-US21353.
XX
PR 29-JUL-1997; 97US-0902615.
PR 14-NOV-1996; 96US-0749051.
PR 01-APR-1997; 97US-0831309.
PR 01-APR-1997; 97US-0833457.
PR 01-APR-1997; 97US-0834705.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX
XX WPI: 1998-297855/26.
DR N-PSDB; AAV52012.
XX
XX Helicobacter polynucleotide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
XX
PS Claim 1; Page 86-88; 362pp; English.
XX
XX This claimed Helicobacter pylori polypeptide, designated GHPO 107,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC AAW71474-W71558) are claimed, as well as isolated polynucleotides
CC (see AAV52009-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transformed or transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention and which can be
CC used to prevent or treat Helicobacter infection; therapeutic and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or monospecific antibodies; methods for
CC detecting the presence of Helicobacter in samples using e.g.
CC the polypeptides or monospecific antibodies; and methods for

CC purifying the polypeptides by antibody-based affinity
CC chromatography.
XX
SQ Sequence 847 AA;
Query Match 5.4%; Score 172.5; DB 19; Length 847;
Best Local Similarity 20.0%; Pred. No. 4-5e-05;
Matches 163; Conservative 91; Mismatches 230; Indels 333; Gaps 37;
QY 76 TQOQEVLDKQQTGF-----LAEAPDNVKTMLRSGYFSSKVS-L-TK--DGAYT 123
Db 81 TEKERDGL-KSOMGKIKGDTFDEQKLEHARTALTEGOGYGVSVVEVTEKVEGALL 139
QY 124 VHIITGPRTKIANVGVAI-LGDILSDGNLAEEYRNALENWQOPVSGDFDQDSW----- 175
Db 140 IVFD-----VNRGDSIYIKQSIYEGS-AKLKRRMIESLSANKORDPFGWMLNDCKL 191
QY 176 -----ENSKTSVLGAVTRKY----- 191
Db 192 RLDQLEYSMRIQDVMRGRGLDAHISSPFLKTFDSTHDAKLHYKVKEGIQIRISDILIE 251
QY 192 -----PLAKL-----GNTAAVNPDTATVDLNV 214
Db 252 IDNPVPLKTLKALKVKRKDVFNIEHLRADAQILKTEADKGYAFVVKPDLDRKNG 311
QY 215 VV-----DSGRPIAFGDFEITGTORYPEQIVSGLARFQPGTPDLDLLDFOQALEONG 268
Db 312 LVKVTYREVGDMYINDVIISGNQRTSDRIIRRELLIQLQPKYINLTJLRNSENLSRLIG 371
QY 269 HYSGASVQADFRLQGRVPVKVSVTEVKRHKLETGIRLDSYGLGKLTAYDYNNLFNKG 328
Db 372 FFS--KVKIEEKRVNSSLMDLLVSVESGRTGQLQFGYGSYGLMLNGSVSENLFGTG 429
QY 329 YIGSVWVD-----MDKYETTLAG--ISOPR-----NYR----- 355
Db 430 QSMSLYANIATGGGSGYPMKPGAGRMFAGNLSLTNPRIFDMSWYSSTINLYADYRISYOY 489
QY 356 -----GNYWTSNVSYNRSTTQNLKRAFSGGIWYVRDRAGID----- 392
Db 490 IQQGGGFGVNVGRMLGNRTHSVGLYNLNVTKLL--GFSSPL-YNRYTSSVNEVVSPPQC 545
QY 393 ---ARLGAEEFLAEGR-----KIPGS-----DIDLGN- 415
Db 546 STPASVIINRUSGGKTPLOPESCSSPGAITTSPETIRGIMDRDYHTPTITSFTLDVSYDNT 605
QY 416 -----SHATM--LTASWKRQLLNVLHPENGHYLDGKIGTT--LGTF----- 453
Db 606 DDYVYFRNGVIFSSYATMSGLPSS-----GTLNSWNG--LGNVNRNTKVYKKAAYHH 656
QY 454 LSTALIRTSARAGYFFTPENKKLGTFIIRGQAGYTVARDNAD--VPSGLMFRSGGASSV 511
Db 657 LQKYLIDILIAH-----FRYNTDDYLPNLNSTFYMGVGTIV 699
QY 512 RGYELDSIGLAGPNSVLP--ERALLVGS-----L 539
Db 700 RGFR-----NGSVTPKDFEGLMLGGDIFTTASTELSYGVVLKAAKMLAWFFDFGFL 750
QY 540 EYQLP-----PTRLTSCAVFHDMDAAANFKRMKLLKHGSLGVRWFSLAP--PSFDA 591
Db 751 TKETPRGSFFYNAPVTNTANFKDYVIGAGFERATWKRSTGQIIEWISPMGPLVLIFPIA 810
QY 592 YGH-----SDKKIR-----WHISLGTFR 609
Db 811 FPNQMGDGNKKCKGLCFNPNDYDTQHFEEFSMGTRF 847
RESULT 35
ID AAB46311 standard; Protein; 916 AA.
XX AAB46311;
AC AAB46311;
XX
DT 05-APR-2001 (first entry)

H. pylori HPS120 protein.
 Microbial infection; antibacterial; Helicobacter pylori Infection;
 vaccine; screening.
 Helicobacter pylori.
 W0200073502-A2.
 07-DEC-2000.
 31-MAY-2000; 2000WO-EP05024.
 31-MAY-1999; 99DE-1024965.
 17-JUN-1999; 99DE-1027740.
 21-JUL-1999; 99DE-1034029.
 (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 (CREA-) CREATOGEN GMBH.
 Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF.
 WPI; 2001-049948/06.
 N-PSDB; AAF25588.
 Preparing an agent for diagnosis or control of microbial infection,
 useful particularly against Helicobacter, based on identification of
 essential genes in defective mutants -
 Claim 37; Page 236-239; 366pp; German.
 This invention describes a novel preparation of an agent (A) for
 detection, prevention and/or treatment of microbial infection by:
 (i) identifying essential genes (I) and corresponding polypeptides
 (II); (ii) identifying compounds that are directed against (ii) and
 inactivate the microbe; (iii) testing these for suitability for use; and
 (iv) formulating selected (A). Identifying essential genes (i) comprise
 preparation of gene-deficient microorganisms by conditional antisense
 inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),
 then determining viability and/or survival of the deficient organisms.
 The products of the invention have antibacterial activity. (A) (which may
 be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 fragments or an inhibitor of (Ia)) are particularly used for diagnosis,
 treatment or prevention of infection by Helicobacter pylori. Particularly
 (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 identifies essential genes, including those that have homologs in other
 species, so identified (A) should have a broad spectrum of activity. Many
 gene-deficient cells can be screened quickly, in an automated process,
 and the identified genes can be used for screening without purification.

Query Match 5.4%; Score 172.5; DB 22; Length 916;
 Best Local Similarity 20.0%; Pred. No. 5.1e-05;
 Matches 163; Conservative 91; Mismatches 230; Indels 333; Gaps 37;
 QY 76 TQOEEVLDKEOTGF-----LAEAPDNVKTMLRSKGYFSSKVSU-TEK--DGAYT 123
 Db 150 TERKDLG-KSOMGKKGDTFDEQLEHAKTALTEGGQYGVVVEVTEKVSIGALL 208
 QY 124 VHTPGPRTKIANVGVAI-LGDLSDGNLAERYNALENWQPPGVSDFDQDSW----- 175
 Db 209 IVPD-----VNRGDSIVIKQISIEGS-AKLKRMRIESLSANKQRDFMGMGLNDGKL 260
 QY 176 -----ENSKTSVLGAVTRKY----- 191
 Db 261 RLQLEYDSMRLODYMYRRGYLDHAHSFFLKTDFSTHDAKLHKYKKEGTQYRISILLIE 320
 QY 192 -----PLAKL-----GNRAAVNPDPATVVDLNV 214
 Db 321 IDNPVVPLKLEKALKVRKRVNTEHLRADAQILKTEIADKGYAFVAVKPDLDKDKNG 380

QY 215 VV-----DSGRPIAFGDFEITGTORYPEQIVSGLARFQPGTYPYDLDLLDFOQALEQNG 268
 Db 381 LKVVIVIEVGMVYINDVILSGNORTSDRIIRRELLGPKDKYNLTJTKRNSENSURRLG 440
 QY 269 HYSQASVQADFRLQCDRVPVKVSTEVKRHKLETGRLDSEYGLGKIADYDYNLNFNG 328
 Db 441 FPS--KVKIEKRVNSSLMDLLVSVEEGRTGQLQGLGYSYGLGLMNGSVSRNLEFGT 498
 QY 329 YIGSVVND-----MKYETTLAAG---ISQPR-----NVR----- 355
 Db 499 QSMSLYANITATGGRSYPCMPKAGRMFAGNLSITNPRIFDSWYSSTINLYADYRISYOY 558
 QY 356 -----CNYWTSNVSYNRSTTNQLEKRAFSGGIWTVRDRAGID----- 392
 Db 559 IQGGGGFVNVGRMLGNRTHVSLGYNLNVTKLL--GFSSPL--YNYRYSVSNVSWSPROC 614
 QY 393 ---ARLCAEFLAEGR-----KIPGS-----DIDLGN- 415
 Db 615 STPASVILNRLSGKTPLOPESCSPPGAIITTSPIRGIWDRDYHTPITSSFTLDVSYDNT 674
 QY 416 -----SHATM--LTASMKRQLLNVLHPENGHYLDCKIGCTT--LGTF----- 453
 Db 675 DDYFPRNGVIESSYATKSLPSS-----GTLNSNG--LGNVNRNTKVKCKFAAYIHH 725
 QY 454 LSTALIRTSARAGYFFTPENKKLGTFFIRQOAGYTVARDNAD--VPSGLMFRSGGASSY 511
 Db 726 LQYLLIDLIAR-----FKTGGY-IFRYNTDDYLPNLTSTFYMGGVTV 768
 QY 512 RGYELDSIGLAGPNSVLP--ERALLVGS-----WHISLGRF 609
 Db 769 RGR-----NGSVTPKDEFLWLGDDGIFTASTELSVGLKAAKMRAMWFFDFGFL 819
 QY 540 EYOLP-----FTRTLGAVFHDMDGDAANFKRMKHLKHSGLGVNRFSLAP--FSFDIA 591
 Db 820 TFKTPTRGSFFNAPVTTFANFDYGVIGAGFERATWRASTGLQIEWISPMGPLVLFFPIA 879
 QY 592 YGH-----SDKKIR-----WHISLGRF 609
 Db 880 FFQWGDGNGKCKKGLCFNPNDMDYTQHFEPFSGMTRF 916
 RESULT 36
 AAY69362
 ID AAY69362 standard; Protein; 787 AA.
 AC AAY69362;
 DT 19-JUN-2000 (first entry)
 XX Amino acid sequence of the CPN100111 polypeptide.
 XX CPN100111; Chlamydia Infection; immune response; vaccine.
 XX Chlamydia pneumoniae.
 PH Key Location/Qualifiers
 FT Peptide 1..85
 FT /note= "signal peptide"
 XX W0200011183-A2.
 XX
 PD 02-MAR-2000.
 XX
 PF 18-AUG-1999; 99WO-IB01449.
 XX
 PR 20-AUG-1998; 98US-0097187.
 PR 20-AUG-1998; 98US-0097188.
 PR 20-AUG-1998; 98US-0097189.
 PR 20-AUG-1998; 98US-0097190.
 PR 20-AUG-1998; 98US-0097195.
 PR 20-AUG-1998; 98US-0097196.
 PR 20-AUG-1998; 98US-0097197.

QY 151 LAEYRNALENKQPV--GSDFDOD-----SWENSKTSVLGAVTRKGYPLAKLGNTRAAYN 204
 Db 200 LIDVLA-INKDNINPLSRADRYTQEKLVTSLENLRAKYLNA-----GFVREIKDAKLIN 254
 QY 205 PDATVLDLVVVDSGRPTAFGDFEITGTQRYPEQIVSGLARFOGTPYDLDLLDFQOAL 264
 Db 255 EDKNRIFVEISLHEGEQYRFGQTQFLGNTVYTOAELEALLKFAEGEFS-----QAML 307
 QY 265 EQ-----NGHY-----SGASVQADDFRLOGDRVPVKV-----SV 293
 Db 308 EQTTNNISTKFGDDGYIAQIRPVTRINDESRVVDVEYIDPVHPVYVRINFTGNFKTQ 367
 QY 294 TEVKKHLETGIRLSEYGLGKIAIDYNNLFNKG-----IGSVV 334
 Db 368 DEVLRRMR---OLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNPSDQVDVNFV 424
 QY 335 WMDKYETTLLAAGISO-----PRNYRGNYWTSNVSYNRSTTONLEKRAFSGGIW 383
 Db 425 EQPSSSTIAAGYSQSGGVTFQFDVSONNFMGCKHVNASFSRSETREVYSLGWTNPYF 484
 QY 384 -----YVRDRAG-----ID--ARLGAFFLAEGRKIPG 408
 Db 485 TVNGVSQSLSGYRKYTKYDNKNISNYVLDYSGSLSYGYPIDENQRISEGLNADNTKLHG 544
 QY 409 SDIDLGNSHATMLTA-----SNKROLLNNVLHPENG--H 440
 Db 545 GRF-MGISNVKQLMADGGCKIOVDNNGIPDFKHDTTYTNAILGNWYSSLDPRVPTQGM 603
 QY 441 YLDGKTGTTGLTFLSTALIRTSARAGYFFTPENKKGCTFIIRGOA--GYTVARDNADVP 498
 Db 604 SYDLTVG-----FGDKTH--QKVVYQGYNIYRPFIKK--SVLRGYAKLGY-----GNNLP 648
 QY 499 SGLMFRSGCASSVRGYELDSIGLAGPN-----GSLVPERALLVGSLEYOL 543
 Db 649 FYENFYAGGYGVRGYDQSSL--GPRSQAYLTARRGQOQTLGVEVGGNALLATFGSELL 705
 QY 544 PF-----TRTL-----SGAVFHDG-----DAAANPKR-----566
 Db 706 PLFPKGDWIDQVRPVIFIEGGQVFTTGMCKQOTIDLTFQFKDQATAEQNAKANRPLLTQ 765
 QY 567 -MKLKHGSLGVRWFSPPLAPFSFDTA 591
 Db 766 DKOLRYSAGVGATWVTPICPLSISYA 791

RESULT 38

AA444391
 ID AAY44391 standard; Protein; 813 AA.
 AC AAY44391;
 XX
 XX
 XX 14-MAR-2000 (first entry)
 DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
 XX
 XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
 KW diagnosis; bacterial infection.
 XX
 XX Moraxella catarrhalis.
 OS
 XX
 XX W09963093-A2.
 XX
 XX 09-DEC-1999.
 XX
 XX 31-MAY-1999; 99WO-EP03822.
 XX
 XX 03-JUN-1998; 98GB-0011945.
 PR 08-MAR-1999; 99GB-0005304.
 XX
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA
 XX
 XX Vlnals-Bassols C;
 PI
 XX

DR WPI: 2000-105700/09.
 DR N-PSDB; AA229551.
 XX
 PT Novel BASB027 polynucleotide and polypeptides from Moraxella
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis
 PT media
 XX
 PS Claim 1; Page 102-104; 109pp; English.
 PS
 CC The present sequence is BASB027 polypeptide, which shows significant
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
 CC polypeptide can be used for diagnosis and staging of disease, determining
 CC susceptibility to a disease and to prepare medicaments for treating M.
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
 CC used as probe for screening of genetic mutations, serotype, taxonomic
 CC classification or identification. BASB027 agonists, antagonists and
 CC antibodies may be used to prevent and/or treat bacterial infections.
 XX
 XX Sequence 813 AA;
 Query Match 5.2%; Score 165.5; DB 21; Length 813;
 Best Local Similarity 19.4%; Pred. No. 0.00016;
 Matches 145; Conservative 96; Mismatches 260; Indels 245; Gaps 36;
 QY 55 VRIDTQDSIKDMVEHPLITQ---QOEVLKDE--QTGF-----LAEAPDNV 99
 Db 82 VOYHOGEGRIYQVTER-PLIAEINFEGRNLPKREGLOEGLKNAGLAVGOPLQAQTVQMI 140
 QY 100 KTLNR---SKGYFSKSVSL--TEKDG---AYTVHITGPRTKIANVGVAILDGLDILDCGN 150
 Db 141 ETELNTQYISQGYNYTEITVKTQMLDGNRVKLDMTFAEGKPARVVDINI-IGNQHFSDAD 199
 QY 151 LAEYRNALENKQPV--GSDFDOD-----SWENSKTSVLGAVTRKGYPLAKLGNTRAAYN 204
 Db 200 LIDVLA-INKDNINPLSRADRYTQEKLVTSLENLRAKYLNA-----GFVREIKDAKLIN 254
 QY 205 PDATVLDLVVVDSGRPTAFGDFEITGTQRYPEQIVSGLARFOGTPYDLDLLDFQOAL 264
 Db 255 EDKNRIFVEISLHEGEQYRFGQTQFLGNTVYTOAELEALLKFAEGEFS-----QAML 307
 QY 265 EQ-----NGHY-----SGASVQADDFRLOGDRVPVKV-----SV 293
 Db 308 EQTTNNISTKFGDDGYIAQIRPVTRINDESRVVDVEYIDPVHPVYVRINFTGNFKTQ 367
 QY 294 TEVKKHLETGIRLSEYGLGKIAIDYNNLFNKG-----IGSVV 334
 Db 368 DEVLRRMR---OLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNPSDQVDVNFV 424
 QY 335 WMDKYETTLLAAGISO-----PRNYRGNYWTSNVSYNRSTTONLEKRAFSGGIW 383
 Db 425 EQPSSSTIAAGYSQSGGVTFQFDVSONNFMGCKHVNASFSRSETREVYSLGWTNPYF 484
 QY 384 -----YVRDRAG-----ID--ARLGAFFLAEGRKIPG 408
 Db 485 TVNGVSQSLSGYRKYTKYDNKNISNYVLDYSGSLSYGYPIDENQRISEGLNADNTKLHG 544
 QY 409 SDIDLGNSHATMLTA-----SNKROLLNNVLHPENG--H 440
 Db 545 GRF-MGISNVKQLMADGGCKIOVDNNGIPDFKHDTTYTNAILGNWYSSLDPRVPTQGM 603
 QY 441 YLDGKTGTTGLTFLSTALIRTSARAGYFFTPENKKGCTFIIRGOA--GYTVARDNADVP 498
 Db 604 SYDLTVG-----FGDKTH--QKVVYQGYNIYRPFIKK--SVLRGYAKLGY-----GNNLP 648
 QY 499 SGLMFRSGCASSVRGYELDSIGLAGPN-----GSLVPERALLVGSLEYOL 543
 Db 649 FYENFYAGGYGVRGYDQSSL--GPRSQAYLTARRGQOQTLGVEVGGNALLATFGSELL 705
 QY 544 PF-----TRTL-----SGAVFHDG-----DAAANPKR-----566
 Db 706 PLFPKGDWIDQVRPVIFIEGGQVFTTGMCKQOTIDLTFQFKDQATAEQNAKANRPLLTQ 765

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant products of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX	Sequence	371 AA;
SQ	Query Match	5.1%; Score 160.5; DB 22; Length 371;
	Best Local Similarity	23.9%; Pred. No. 0.00013;
	Matches	77; Conservative 27; Mismatches 103; Indels 115; Gaps 13;
QY	355 RGNW-----TSNVSYNRSTTONLEKRAFSGGIWYVDRAGIDARLGAEFLAEGRKI	406
Db	16 RGVYWDPNSDLCPTLDFSYKMPLLN-----PLEQYLLVQG-GF	53
QY	407 PGSDIDLGNSHATMLTAS-----WKROLLNNVLPENGHYLDGKIGTITGTFLSSTA	458
Db	54 KRTDLNDESDSTTLVASRYWDLSSGWQRAI---NLRWSLDHFTQGEINTTMTLFPQVM	110
QY	459 LIRTSARAGYF-----FTPENKKLOT-----FIIRGOA	486
Db	111 ISRTRSGGLMPTWGDQSQRYSIDYSNTAWGSDVDFSVFQAQNVWIRTYDRHRFVTRGTL	170
QY	487 GYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNCGLPERALLVGSLEYQLPPT	546
Db	171 GWIETGDFKVPDDLRFPAAGDRSIRGYQYKSGPKLPNGDLKGASKLITGSLEYQYNYT	230
QY	547 -----RTLSGAVF-----HDMGDAANFK---RMKLK-HGSG	574
Db	231 GKCTKWQGEWLAPEGEVRLAPGATLATPELVASCSTEGNLGLAANFHAELRARLPWHGGA	290
QY	575 LGVR-----WFSPLAPFSFDI	590
Db	291 MKPRPVHLNTW----EGFYFDL	308

Search completed: November 9, 2002, 01:18:45
Job time : 44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: November 9, 2002, 01:24:34 ; Search time 57 Seconds
(without alignments)
3276.599 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGHSKIRWHISIGTRF 609

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=vlh
-Q=/cgn2_6/ptodata/1/ina/5A_COMB.seq:
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669_ECGN_1.1.19 -runat.05112002.110008.7750 -NCPD=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:
5: /cgn2_6/ptodata/1/ina/PCUS_COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213.5	6.7	2989	3	US-08-433-522A-9 Sequence 9, Appli
2	213.5	6.7	2989	3	US-09-135-166-9 Sequence 9, Appli
3	213.5	6.7	2989	4	US-08-942-046-9 Sequence 9, Appli
4	210.5	6.6	2974	3	US-08-433-522A-7 Sequence 7, Appli
5	210.5	6.6	2974	3	US-09-135-166-7 Sequence 7, Appli
6	210.5	6.6	2974	4	US-08-942-046-7 Sequence 7, Appli
7	209.5	6.6	2949	3	US-08-433-522A-1 Sequence 1, Appli
8	209.5	6.6	2949	3	US-09-135-166-1 Sequence 1, Appli
9	209.5	6.6	2949	4	US-08-942-046-1 Sequence 1, Appli
10	209.5	6.6	2950	3	US-08-433-522A-5 Sequence 5, Appli
11	209.5	6.6	2950	3	US-09-135-166-5 Sequence 5, Appli
12	209.5	6.6	2950	4	US-08-942-046-5 Sequence 5, Appli

13	209.5	6.6	2984	3	US-08-433-522A-3 Sequence 3, Appli
14	209.5	6.6	2984	3	US-09-135-166-3 Sequence 3, Appli
15	209.5	6.6	2984	4	US-08-942-046-3 Sequence 3, Appli
16	209.5	6.6	2987	3	US-08-433-522A-55 Sequence 55, Appli
17	209.5	6.6	2987	3	US-09-135-166-55 Sequence 55, Appli
18	209.5	6.6	2987	4	US-08-942-046-55 Sequence 55, Appli
19	134.5	4.2	1697	4	US-09-346-408-5 Sequence 5, Appli
20	121.5	3.7	2280	4	US-09-346-408-3 Sequence 3, Appli
21	117	3.7	1890	4	US-09-346-408-7 Patent No. 5223391
22	114.5	3.6	1634	6	5223391-4 Patent No. 5223391
23	114	3.6	9171	1	US-08-038-682-5 Sequence 5, Appli
24	114	3.6	9171	1	US-08-302-832-5 Sequence 5, Appli
25	114	3.6	9171	2	US-08-530-198-5 Sequence 5, Appli
26	114	3.6	9171	2	US-08-469-880-5 Sequence 5, Appli
27	114	3.6	9171	2	US-08-728-470-5 Sequence 5, Appli
28	114	3.6	9171	2	US-08-617-697-5 Sequence 5, Appli
29	114	3.6	9171	4	US-08-719-641-5 Sequence 5, Appli
30	112.5	3.5	1524	1	US-08-021-601-9 Sequence 9, Appli
31	112.5	3.5	1524	1	US-08-082-849B-9 Sequence 9, Appli
32	112.5	3.5	1524	5	PCT-US94-01624-9 Sequence 9, Appli
33	112	3.5	3286	2	US-08-363-124A-1 Sequence 1, Appli
34	111.5	3.5	1791	1	US-08-565-386-10 Sequence 10, Appli
35	111.5	3.5	19227	3	US-09-090-793-13 Sequence 13, Appli
36	111.5	3.5	40138	3	US-09-080-793-12 Sequence 12, Appli
37	111	3.5	5100	1	US-08-164-292B-15 Sequence 15, Appli
38	111	3.5	5100	1	US-08-164-292B-17 Sequence 17, Appli
39	111	3.5	5100	1	US-08-164-292B-19 Sequence 19, Appli
40	111	3.5	5100	1	US-08-164-292B-21 Sequence 21, Appli
41	111	3.5	5100	1	US-08-164-292B-23 Sequence 23, Appli
42	111	3.5	5100	1	US-08-164-292B-25 Sequence 25, Appli
43	111	3.5	5100	3	US-08-845-623-15 Sequence 15, Appli
44	111	3.5	5100	3	US-08-845-623-17 Sequence 17, Appli
45	111	3.5	5100	3	US-08-845-623-19 Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-08-433-522A-9
; Sequence 9, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163

US-09-135-166-9
 : Sequence 9, Application US/09135166
 : Patent No. 6083743
 : GENERAL INFORMATION:
 : APPLICANT: CHONG, Pele
 : APPLICANT: THOMAS, Wayne
 : APPLICANT: YANG, Yan ping
 : APPLICANT: LOOSMORE, Sheena
 : APPLICANT: SIA, Dwo Yuan Charles
 : APPLICANT: KLEIN, Michel
 : TITLE OF INVENTION: HAMOPHILUS OUTER MEMBRANE PROTEIN
 : NUMBER OF SEQUENCES: 55
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Sim & McBurney
 : STREET: 6TH Floor, 330 University Avenue

```

1 CITY: Toronto
2 STATE: Ontario
3 COUNTRY: Canada
4 ZIP: #5G 1R7
5
6 COMPUTER READABLE FORM:
7
8 MEDIUM TYPE: Floppy disk
9
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: PatentIn Release #1.0, Version #1.25
13
14 CURRENT APPLICATION DATA:
15
16 APPLICATION NUMBER: US/09/135.166
17

```

CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-829 MTS:jd
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 390..2768
US-09-135-166-9

Alignment Scores:	
pred. No.:	6,28e-13
Score:	23.50
Length:	2389
Matches:	127
Conservative:	85
Mismatches:	254
Best Local Similarity:	20.99%
Query Match:	6.73%
Indels:	139
Gaps:	24
DB:	3

US-09-857-669-2 (1-609) x US-09-135-166-9 (1-2989)

Qy	105	SerLysGlyTyrPheSerSerLysValSerLeuThr	-----GluLys	118
		::: :::: :::: :::: ::::		
Db	1071	AAATAATGGCTATGCCAAACCAATCATAAACGGATGTT	CAGCTAAATCATGAAAA	1130
		::: :::: :::: :::: ::::		
Qy	119	AspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys	-----	133
		: :::: :::		
Db	1131	ACAAAAGTAAATGTAAACCATGTATGTAATGAAGGTTTACAGTACACCTCGTAGTGCA	1190	
Qy	134	-----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGly	149	
		: :::: ::::		
Db	1191	CGCATATTAGGTAATCTGGGAGGTATGTCTCCGAGCTTGAACCTTTACTTTACTTCACGATTA	1250	

Qy	150	AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp	169
Db	1251	CAITTAATGAATGACTTTCCGCGGTAGT	1277
Qy	170	PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys	189
Db	1278	---GATATTGCAGATGTAGAAAATSCAATTAAAGCAAAACTTGGG---GAACGA	1325
Qy	190	GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp	206
Db	1326	GGTTAC-----GGAACACACAGTAATAATCTGTACCTGATTTTGACGAT	1370
Qy	207	---ThrAlaThrValAspLeuAsnValValAlaAspSerGlyArgProIleAlaPheGly	225
Db	1371	GCAAAATAAACATACCGATAACCTTGTGTGTGATGCTGGCAGCGTTTAACTGTCGCG	1430
Qy	226	AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg	245
Db	1431	CAACTTCGCTTTGAAGAAATACCGTTCCTGCTGATGACTTCTACGTCAGGAATCGGA	1490
Qy	246	PheGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGlnAlaLeuGlu	265
Db	1491	CAACAAGCAAGCACTTTGGTATAATTACCAATTAGTTAGTTAGGAAATATTCGCTAGAT	1550
Qy	266	GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly	284
Db	1551	CGTACAGGTTCTTCGAA-----ACAGTTGAAACCGAATGATCTCATCGGTAGGC	1604
Qy	285	---AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr	303
Db	1605	AATCATCAAGTCGATGCTGTATATAAAGTCAAGAAGCACTAACCGGTAGTATCAACTTT	1664
Qy	304	GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsn	323
Db	1665	GGTATTGGTTCACGGTACAGAGAGTGGTATCAGTTATCAACAAAGTATTAAACAGATAT	1724
Qy	324	LeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThr	343
Db	1725	TTCTTGGGAACAGGGCGGCAGTAGATATAGCTGGTACGAAAAATGATTTGGTACGAGT	1784
Qy	344	LeuAlaAlaGlyIleSerGlnProArg-----AspTyrArgGlyAsn	357
Db	1785	GTCAATTTGGTTATACCGAACCCATTTTCTAAAGATGGGTAAAGCTTTGGTGGAAAT	1844
Qy	358	TyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArg	376
Db	1845	ATTTCTTTGAAACTACGATACTCTAAAGTGATACATCTCTTAACATATTAACGCGTACG	1904
Qy	377	AlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAlaGlyIleAspAla	393
Db	1905	ACTTATGGAAGTAACTGTACTTTAGTTTCCCTGTAAATGAAATAACTCCTATATGTA	1964
Qy	394	ArgLeuGly-----AlaGluPheLeuAlaGlu-----	402
Db	1965	GGATTAGGCCATACCTATAATAAAATTAGTAATTTGCTCTAGATAATAACCGTAATTTA	2024
Qy	403	-----GlyArgLysIleProGlySerAspIleAspLeuGly	414
Db	2025	TATATTCAATCAATGAATTTAAAGCTAATGCGATTTAAACAATGACTTTGAT-----	2078
Qy	415	AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu	434
Db	2079	-----TTTTCTTTTGGTTGGAACTATPAACAGCGCTTAATAGAGCGCTAT	2120
Qy	435	HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu	454
Db	2121	TTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGGTGGCAGGTTACTATTCCAGGT	2174
Qy	455	SerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsn	474
Db	2175	TCTGATACAAATACTACAAACTAAGTCAGATGTACAGGGTTTCTACCCATTAGACAGA	2234

QY 475 LysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaAraGAspAsn 494
Db 2235 GATCACCGCTGGTGTATCTGCAAAACCATCTCGAGATATGCAATGGTTTGGAAAC 2294
QY 495 AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyr 514
Db 2295 AAGCGTTTACCTTCTATCAAACTTATACACGGGTGGCATTTGTTTCATTACGCGGTTT 2354
QY 515 GluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2355 GCTTATGGTAGTATT-----GGGCTAATGCAATTTATGCCGAACATGGTAAATGGT 2405
QY 526 -----GlySerValLeuProGluArgAlaLeuValGlySerLeu 539
Db 2406 ACTTTTAAAGATAAGTCTGATGTGATTGGTGAATGCAATCAACATCGGAGTGCA 2465
QY 540 GluTyrGlnLeuPro-----PheThrArgThr----- 548
Db 2466 GAATTATTGTACCAACTCCATTTCTGAGTGATAAAAGCCAAATACAGTCGAACCTCC 2525
QY 549 -----Leu 549
Db 2526 CTATTTGTTGATCGCGCAAGTGTTTGGCAATATAAATGGAATCAGATAAAATGGATT 2585
QY 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLysLeu 569
Db 2586 GAGACGAGCTCTTGAAGACTTACCTGATTATGCG-----AAATCAACCGGTATT 2636
QY 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
Db 2637 CGCGCTCTACAGGTGCGGATTCGAATGGCAATCTCTATTGGACCATTTGGTATTTTCT 2696
QY 590 IleAla-----TyrGlyHisSerAspLysLysLysIleArgTrpHisIleSer 604
Db 2697 TATGCTAAACCAATTAATAATATGAAATGATGATGTCGAA---CAGTTCCAAATTAGT 2753
QY 605 LeuGlyThrArgPhe 609
Db 2754 ATTGGGGGCTCTTTC 2768

RESULT 3

US-08-942-046-9
: Sequence 9, Application US/08942046
: Patent No. 6264954
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/942,046
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2889 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 390..2768
: US-08-942-046-9
Alignment Scores:
Pred. No.: 6,28e-13 Length: 2989
Score: 213.50 Matches: 127
Percent Similarity: 35.04% Conservative: 85
Best Local Similarity: 20.99% Mismatches: 254
Query Match: 6.73% Indels: 139
Gaps: 24
US-09-857-669-2 (1-609) x US-08-942-046-9 (1-2989)
QY 105 SerLysGlyTyrPheSerSerLysValSerLeuThr-----GluLys 118
Db 1071 AATAATGGCTATGCCAAAGCAAAATCACTAAACGCGATGTTACAGCTAAATGATGAAAA 1130
QY 119 AspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys----- 133
Db 1131 ACAAAAGTTAATGTAAACCATTTGATTAATGAAGGTTTACAGATGACCTTCGTAGTGA 1190
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGly 149
Db 1191 GCGATTATAGTAATCTGGAGGTATGCTGCCGAGCTTGAACCTTTACTTTCAGCATTA 1250
QY 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuAlaAsnTrpGlnGlnProValGlySerAsp 169
Db 1251 CATTTAATGATACCTTTCGCGGTAGT----- 1277
QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
Db 1278 ---GATATTGCAGATGTAGAAAATGCAATTAAGACAAACTTGGG-----GAACGA 1325
QY 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp----- 206
Db 1326 GGTTC-----GGTAACACACACTAAATCTGTACTGATTTTGACGAT 1370
QY 207 ---ThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGly 225
Db 1371 GCAATATAAACCATTTAGCGATACCTTTGTTGTTGCTGGAGCAGGTTTAACTGTTCCG 1430
QY 226 AspPheGluIleThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaAraG 245
Db 1431 CAACTTCGCTTTCAAGGAATACCGTTCTGCTGATAGTACTTACTCAGGAAATGCGA 1490
QY 246 PheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu 265
Db 1491 CAACAAGAAGAACTTGGTATATAATTCACAATTAGTTAGTAGGAAAAATTCGCTTAGAT 1550
QY 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly--- 284
Db 1551 CGTACAGGTTTCTTCGAA-----ACAGTTGAAACGAATTGATCTATCATGTTAGC 1604
QY 285 ---AspArgValProValLysValSerValThrGluLysArgHisLysLeuGluThr 303
Db 1605 AATGATGAAGTGGATGCTGTATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTT 1664

```

Db 2697 TATGCTAAACCAATAAAAAATATGAAAATGATGATGTCGAA---CAGTTCACAAATTACT 2753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 605 LeuGlyThrArgPhe 609
:||||| |||
Db 2754 ATTGGGGCTCTTTC 2768

RESULT 4
US-08-433-522A-7
: Sequence 7, Application US/08433522A
: Patent No. 6013514
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & Mcburney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/433,522A
: FILING DATE: 12-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2974 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 386..2761
US-08-433-522A-7

Alignment Scores:
Pred. No.: 1-32e-12 Length: 2974
Score: 210.50 Matches: 121
Percent Similarity: 34.88% Conservative: 75
Best Local Similarity: 21.53% Mismatches: 241
Query Match: 6.63% Indels: 125
DB: 3 Gaps: 21

US-09-857-669-2 (1-609) x US-08-433-522A-7 (1-2974)
Qy 134 ileAlaasnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu 151
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1193 ATAGGTAATCTGGGAGGTATGCTGCGGACCTTGAACTTTACTTTCAGCATTTACATTTA 1255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 152 AlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAsp 171
: : : : : |||
Db 1253 AATGATACATCTTCGCGGTAGT-----GAT 1270
: : : : : |||

```

QY 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
Db 1277 ATTGCAGATGTAGAAATGCAATTAAGCAAACTTGGG-----GAACGAGGTATAC 1327
QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
Db 1328 -----GCTAACACACACAGTAATTCCTGATCTTTGACGATGCAAAAT 1372
QY 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 1373 AAACACTTAGCGATAACCTTTGTTGATGCTGGACGAGCTTTACTCTTCACCAACTT 1432
QY 228 GuilleThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 1433 CGCTTTGAAGGAATACCGTTTCTGCTGATAGTACTTTAGCTCAGGAATGCGCAACAA 1492
QY 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
Db 1493 GAAGGAACCTTGGTATATTCACAAATAGTTGAGTTAGGAATAATTCGCTAGATCGTACA 1552
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
Db 1553 GCTTTCTTCGAA-----ACAGTTGAAACCGAATTCATCCTATCAATGGTAGCAATGAT 1606
QY 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
Db 1607 GAAGTGGATGCTGATATATTAAGTCAAGAACGTAACACCGGTAGTATCAACTTTGGTATT 1666
QY 306 ArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
Db 1667 GCTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTCAACAAAGATAATTTCTTG 1726
QY 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrCluThrThrLeuAla 345
Db 1727 GCAACAGGGGGCGGAGTAGTAAAGTACGTTGCGGAAATGTTTCTGATGATGATGAT 1786
QY 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTrp 359
Db 1787 TTGGGTTATACCGAGCCCTATTTTACTAAAGATGGTGAAGTCTTGGTGAATGTTTTC 1846
QY 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGlyLysArgAlaPhe 378
Db 1847 TTTGAAACTAGGATAACTCTAAAGTGATACATCTCTACTATAGCGTACGACTTAT 1906
QY 379 SerGlyGlyIle-----TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
Db 1907 GGAAGTAATGTACTTTAGTGTTCCTGTAATGAAATAACTCTCTATTATTAGGATTA 1966
QY 396 Gly-----AlaGluPheLeuAlaGlu----- 402
Db 1967 GCCCATACCTATAATAAAATAGTAATTTGCTCTAGATATAACCGTAATTTATATT 2026
QY 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
Db 2027 CAATCAATCAAAATTAAGGTAATGGCAATTAACAAACAACTTTGAT----- 2074
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 2075 -----TTTTCTTTGGTGAACATAACAGCGCTTAATAGAGCGTATTTCCTCA 2122
QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 2123 ACTAAGGG-----GTTAAGACGAAGTCTTGGTGGACGAGTTACAAATTCAGGTTCTGAT 2176
QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 2177 ACAAACTACTACAACTAGTCAGAGTGTACAGGGTTTCTACCCATTAGACAGATCATC 2236
QY 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAsp 496
Db 2237 CTCTCGGGTGTATCTGCAAAGACATCTGCAGGATATGCAATGGTTTTCGAAACCAAGCGT 2296
QY 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516

Db 2297 TTACCGTTCTATCAAACTTATACAGCGGTGGCATTTGGTTTCATTACCGGTTTTTGTATT 2356
QY 517 AspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2357 GGTAGCAATT-----GGCCCTAACCCAATTTATCAAGGTCAAAATAATAATTAAT 2407
QY 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
Db 2408 AGATAAGTTCTGATGTGATTTGGTGAATGCAATCGCTACAGTACGACGAGTAAATT 2467
QY 543 LeuPro-----PheThrArgThr----- 548
Db 2468 GTGCCAACTCCATTTGTGAGTGATAAGAGTCAAAATACAGTCCGAACCTCCCTATTGTT 2527
QY 549 -----LeuSerGlyAla 552
Db 2528 GATGCGGCAAGTGTGGAATACTAAATGGAATCAGATAAAATGGATTAGAGACCAAT 2587
QY 553 ValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeuLysHisGly 572
Db 2588 GTCTTGAAAGACTTACCCTGATTATGCG-----AAATCAAGCGGTACTCGCGCTCT 2638
QY 573 SerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla--- 591
Db 2639 ACAGGTGCGGATTCCTCAATGCAATCTCTAGTGGACCACTGCTATTTCTTATGCTAAA 2698
QY 592 -----TyrGlyHisSerAspLysIleArgTrpHisIleSerLeuGlyThr 607
Db 2699 CCAATTAATAAATATGAAATGATGATGTCGAA---CAGTTCCAATTTAGTATTGGGGT 2755
QY 608 ArgPhe 609
Db 2756 TCTTTC 2761
RESULT 5
US-09-135-166-7
; Sequence 7, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, DWO Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1K7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2974 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 386..2761
US-09-135-166-7

Alignment Scores:

Pred. No.: 1.32e-12 Length: 2974
Score: 210.50 Matches: 121
Percent Similarity: 34.88% Conservative: 75
Best Local Similarity: 21.53% Mismatches: 241
Query Match: 6.63% Indels: 125
DB: 3 Gaps: 21

US-09-857-669-2 (1-609) x US-09-135-166-7 (1-2974)

QY 134 IleAlaAsnValGly-----ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeu 151
DB 1193 ATAGTAACTGGGAGGTATGCTGCGGAGCTTGAACCTTTTACTTTTCAGCATATACATTTA 1252
QY 152 AlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAsp 171
DB 1253 AATGATACTTTCCGCGGTAGT-----GAT 1276
QY 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyr 191
DB 1277 ATTGCAGATGTAGAAATGCAANTTAAGCAAACTTGGG-----GAACGAGGTTAC 1327
QY 192 ProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp-----Thr 207
DB 1328 -----GGTAACACAACTAAATTCGTACCTGATTTTACGTAGCAATGCAAAAT 1372
QY 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
DB 1373 AAACATTAGCGATACCTTTGTTGTCCTGGGAGCAGCTTTACTCTTCACCAACTT 1432
QY 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
DB 1433 CGCTTTGAAGGAATACCGTTTCTGCTGATAGTACTTTACGTAGCAAAATGCGCAACAA 1492
QY 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsn 267
DB 1493 GAAGGAACCTTGGTATATTCAAATTTAGTTAGCTTAGGAAATATCGCTTAGATCGTACA 1552
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly-----Asp 285
DB 1553 GCTTTCTTCGAA-----ACAGTTGAAACCGGAATGATGCTCAATCAATGGTAGCAATGAT 1606
QY 286 ArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIle 305
DB 1607 GAAGTGGATGCGTATATAAGTCAAGAAGCGTAACACGGGTAGTATCAACTTTGGTATT 1666
QY 306 ArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTrpAsnLeuPhe 325
DB 1667 GGTTCAGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTCAACAAGATAATTTCTTGT 1726
QY 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAla 345
DB 1727 GGAACAGGGCGGCGAGTAGTAACTAGCTGGTGAAGAAATGATTATGGTACGAGGTGCAAT 1786
QY 346 AlaGlyIleSerGlnProArg-----AsnTyrArgGlyAsnTyrTrp 359
DB 1787 TTGGGTATACCGAGCCCTATTTTACTAAAGATGGTGAAGTCTTGGTGAATAATGTTTC 1846
QY 360 ThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsnLeuGluLysArgAlaPhe 378

DB 1847 TTTGAAAACACGATAACTCTAAAGGTGATACATCCTCTAACTATAGCGTAGCGACTTAT 1906
QY 379 SerGlyGlyIle-----TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu 395
DB 1907 GGAAGTAATGTTACTTTAGGTTTCCCTGTAAATGAAATAAAGTAACTCTATTATCTAGGATTA 1966
QY 396 Gly-----AlaGluPheLeuAlaGlu----- 402
DB 1967 GCCCATACCTATATAAATAATAGTAACTTTGCTCTAGATATAACCGTATATTTATATATT 2026
QY 403 -----GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSer 416
DB 2027 CAATCAATGAAATTTAAAGGTAATGCGCATTAACAAACAATGACTTTGAT----- 2074
QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
DB 2075 -----TTTTCTTTGGTGGAACTATAACAGCCTTTAATAGAGGCTATTTCCCA 2122
QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
DB 2123 ACTAAAGGG-----GTTAAAGCAAGTCTTGGTGGAGGAGTTACAATTCACAGTTCTGAT 2176
QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
DB 2177 ACAAAATACACTACAAGTAAAGTCAGATGTACAGGGTTTCTACCCATTAGCAGAGATCAC 2236
QY 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaAlaArgAspAsnAlaAsp 496
DB 2237 CTCCTGGTGTATCTGCAAAAGCATCTCGAGGATATCAAAATGGTTTGGAAACAGCGT 2296
QY 497 ValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeu 516
DB 2297 TTACCGTCTCATCAACTTTATACAGCGGTGGCATGTTGTTTCATTACCGGTTTTCGCTTAT 2356
QY 517 AspSerIleGlyLeuAlaGlyProAsn----- 525
DB 2357 GGTAGCATT-----GGGCTTAAGCAATTTATCAAGGTCAAAATAAATAATTTAAT 2407
QY 526 -----GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGln 542
DB 2408 AGATAAGTTCTGATGTGATTTGGTGGTAAATGCAATCGCTACAGCTAGCGCAGAGTTAATT 2467
QY 543 LeuPro-----PheThrArgThr----- 548
DB 2468 GTGCCAACCTCCATTTGTGAGTGATAAGAGTCAAAATACAGTCCGAAACCTCCCTATTGTT 2527
QY 549 -----LeuSerGlyAla 552
DB 2528 GATGCGGCAAGTGTGTTGGAATPACTAAATGGAATCAGATATAAATGGATTAGAGAGCAAT 2587
QY 553 ValPheHisAspMetGlyAspAlaAlaAlaAsnPheLysArgMetLysLysLeuLysGly 572
DB 2588 GTCTTGAAGAGACTTACCGGATTTATGGC-----AAATCAAGCGCTACTCGCGCTCT 2638
QY 573 SerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAspIleAla--- 591
DB 2639 ACAGGTGTGGATTCCTCAATGGCAATCTCTAGTGGACCGCTGTTATTTCTTATGCTAAA 2698
QY 592 -----TyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyThr 607
DB 2699 CCAATTAATAAATAATGAAATGATGATGTCGAA---CAGTTCCAATTTAGTATTTGGGTT 2755
QY 608 ArgPhe 609
DB 2756 TCTTTC 2761

RESULT 6

US-08-942-046-7
; Sequence 7, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele

Db 804 AATGATGAAAGAAACAAAGTTAAATGTAACCAATTGATGTAATGAAGTTTACAGTATGAC 863
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 864 CTTTCGAGTGCACGCAATTAAGTAATCGGAGGTATGTCTCCGAGCTTGAACCTTTA 923
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
Db 924 CTTTCAGCATTACATTAATGATACITTCGCGCGTAGT----- 962
QY 166 ValGlySerAspGlyAsnLeuAlaGluTyrTyrArgAsnLeuGlyAla 185
Db 963 -----GATATTGCAGATGTAGAAAATCAATTAAGCAAAACACTTGA--- 1004
QY 186 ValThrArgGlyGlyTyrProLeuAlaLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1005 -----GAACGCGTTACGGTAGCGCAACGCTAAATTCAGTA-----CCT 1043
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1044 GATTTGATGATGCAAAATAAACATAGCGATAACCTTGTGTGATCGTGACGACGT 1103
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1104 TTAACGTGTTCCGCAACTTCGCTTGAAGGAATACCGTTCTGCTGATACCACTTTACGT 1163
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1164 CAGGAATCGCGCAACAGAGGAACTTGGTATATTAATTCACAATTAAGTTAGTTAGAA 1223
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1224 ATTGCGTTAGATCGTACAGGTTCTTCGAA-----ACATCGAAACCAATTTGATCCT 1277
QY 282 LeuGlnGly-----AspArgValProValValValSerValThrGluValLysArgHis 299
Db 1278 ATCAATGGTGAATGATGAAGTGGATGCTGTATATAAGTCAAGAACGTAACACGGGT 1337
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 1338 AGTATCAACTTTGGTATGTTACGGTACAGAGAGTGTTAGTTATCAAGCAAGTGT 1397
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1398 AACAAGATAATTTCTGGGAACAGGGCGGCAGTAAATATAGTATGCTGTAACGAAAAATGAT 1457
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1458 TATGGTACGAGTCAATTTGGGTTATACCGAGCCCTATTATTAAAGATGGTGAAGT 1517
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer-----ThrThrGlnAsn 372
Db 1518 CTTGGTGGAAATGTTTCTTTGAAACATPACGATAACTCTAAAAGTATACATCCTCTAAC 1577
QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1578 TATAAGCTACGATACGGAAGTAATGTTACTTTAGTTTCCCTCTAAATGAATAATAC 1637
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1638 TCCTATTATGATAGGATTAGGTCATACCTATAATAAATATAGTAACTTTGTCTAGAAATAT 1697
QY 403 -----GlyArgLysIleProGlySerAsp 410
Db 1698 AACCGTAATTTATATATTCATCAATGAATGAATTAAGGTAATGCATTAACAAATGAC 1757
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1758 TTTGAT-----TTTTCTTTTGGTGGAACTATAACACGCTT 1793
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1794 AATAGAGGCTATTTCCCAACTAAAGGG-----GTTAAAGCAAGCTCTGTGGACGAGTT 1847

QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1848 ACTATTCCAGGTTCTGATAACAATACTACAACTAAGTGCAGATGTACAGGTTTCTTAC 1907
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 1908 CCATTAGACAGAGATCACTCTGGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAT 1967
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 1968 GGTTCGAAACAAAGCGTTTACCGTTCTATCAAACTTATACAGCGGTGCGATCGTTCA 2027
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2028 TTACGTGGTGTTCATGTTAGTAGTATT-----GGACCTAACGCAATTTATGCGGAA 2078
QY 526 -----GlySerValLeuProGlu 531
Db 2079 TATGGTAATGGTAGTGGTACTTTTAAAGAGATAAGTTCTGTATGTGATTGGTGGT 2138
QY 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2139 AATGCAATCGCTACAGCTAGCGCAGAGTTAATTTGCCAACTCCATTTGTGACGATAG 2198
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2199 AGCAAAATACCGTCCGAACCTCTTATTGTTGTCGCGCAAGTGTTCGAAATACTAAA 2258
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGGATTAGAGAGCGATGATTAAAGATGCTGATTATGGC 2318
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AAATCAAGCGATTATCGCGCTCTACAGGTGTCGATTCCTCAATGGCAATCTCCTATTGG 2378
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGGTATCTCTATGTCCTCAACCAATTAATAATATGAATGATGATCGAA--- 2435
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCTCAATTTAGTATGGAGGTTCTTTC 2465

RESULT 9
US-08-942-046-1
: Sequence 1, Application US/08942046
: Patent No. 6264954
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/942,046
: FILING DATE:

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/433,522
 FILING DATE: 12-SEP-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: STEWART, Michael I
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2949 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 75..2465
 US-08-942-046-1

Alignment Scores:
 Pred. No.: 1,67e-12 Length: 2949
 Score: 209.50 Matches: 130
 Percent Similarity: 35.85% Conservative: 103
 Best Local Similarity: 20.00% Mismatches: 280
 Query Match: 6.60% Indels: 137
 DB: 4 Gaps: 24

US-09-857-669-2 (1-609) x US-08-942-046-1 (1-2949)

Qy	61	AspSerGluIleLysAspMetValGluGluHsLeuProLeuIleThrGlnGlnGlu	80
Db	624	GAATCTGTTAGTAGCAGTACATTAACAAGACAAATGGAATTAACAACCTGATCTTGGTGG	683
Qy	81	GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys	100
Db	684	AAATTATGGGAAATAAATTCGAAGTCGCGCAATTCGAGAAAGATTTGCACTCAATTCGT	743
Qy	101	ThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeuThr	116
Db	744	GATTATATTAAATAATGGCTATGCCAAAGCACAAATTAATAAAGCGGATGTCAGCTA	803
Qy	117	-----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys	133
Db	804	AATGATCAAAACAAACAACTTAATCTAACCATGATGTAATGAAGGTTTACAGTATGAC	863
Qy	134	-----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle	145
Db	864	CTTCGAGTGCACGCAATTAATGATTAATCGGAGGATGATGTCGCGAGCTTGAACCTTTA	923
Qy	146	LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro	165
Db	924	CTTTCAGCATTTAATGATACATCTTCGCGCGTAGT-----	962
Qy	166	ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla	185
Db	963	-----GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACCTTGA---	1004
Qy	186	ValThrArgLysTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro	205
Db	1005	-----GAACGCGTTTACGGTACGCCAACCGTAAATTCAGTA-----CCT	1043
Qy	206	Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro	221
Db	1044	GATTTTGATGATGCAAAATAAATACATAGCATGACCCCTTTGTTGATGCTGACGACGT	1103
Qy	222	IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer	241
Db	1104	TTAACTGTTCCGCAACTTCGCTTCAAGCAAGAAATACCGTTTCTGCTGATACGACTTTACGT	1163

QY 545 PheThrArgThrLeuSerGlyAlaValAlaPheHisAspMet----- 557
Db 2199 AGCCAAATACGGTCGCAACTCTCTATTGTGTGATGCGCGCAAGTGTGGTAATACTAAA 2258
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2259 TGGAAATCAGATAAAATGATAGAGAGCGATGATTAAAAAGATTGCTGATTATGGC 2318
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2319 AANTCAACCGTATTCGGGCTCTACAGGTGTCGGATTCCAATGGCAATCTCCTATTGGG 2378
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2379 CCATTGGTATTCTCTATGCGCAACCAATATAAAAAATATGAAATGATGATGCGAA--- 2435
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2436 CAGTTCCAATTTAGTATTGGAGGTTCTTTC 2465

RESULT 10

US-08-433-522A-5
Sequence 5, Application US/08433522A
Patent No. 6013514
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: LIAO, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2950 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 334..2724

US-08-433-522A-5
Alignment Scores:
Pred. No.: 1.67e-12 Length: 2950
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280

Query Match: 6.60% Indels: 137
DB: Gaps: 24
US-09-857-669-2 (1-609) x US-08-433-522A-5 (1-2950)
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 883 GAATCTGTAGTACAGTACATTACAAGACAATGGAATTAACAACCTCATTCTTGGTGG 942
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 943 AAATATATGGGAAATAAATGGAAGTGGCGCAATTCGAGAGAAGATTTCAGTCAATTCGT 1002
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1003 GATTATTTAAATAATATGCTATGCCAAACACAAATTAATAACGCGATGTCAGCTA 1062
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1063 AATGATGAAAAACAAAAAGTTAATGTAACCATGATGTAATAAGAGTTTACAGTATGAC 1122
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1123 CTTCTAGTGCAGCGATTATAGGTAATCTGGAGGTATGTCGCCGAGCTTGAACCTTTA 1182
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1183 CTTTCAGCATTAACATTTAAATGATACATTTCCCGCGTAGT----- 1221
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1222 -----GATATTGCAGATGTAGAAAAATGCAATTAACCAAACTTGA--- 1263
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 1264 -----GAACGCGTTACGGTAGCGCAACGCGTAAATTCAGTA-----CCT 1302
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 1303 GATTTTGTATGATGCAATAAACAATAGCGATACCCCTGTTGTTGATCGTGCAGCGCT 1362
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1363 TTAATGTTTCGCCAACTTCGTTTGAAGGAAATACCGTTTCTGCTGATAGCATTTACGT 1422
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1423 CAGGAAATCGCCCAACAAGAGAACTTGGTATAATTCACAATTAAGTTAGTTAGCAAAA 1482
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1483 ATTCGCTTAGATCGTACAGGTTTCTTCGAA-----ACAGTCGAAAACCGAATTGATCCT 1536
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1537 ATCAATGGTAGTAATGATGAAGTGTGCTATATAAAGTCAAGAAACGTAACCGGT 1596
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
Db 1597 AGTATCACTTGTGGTATTGGTTACGGTACAGAGAGTGTATTAGTTATCAAGCAAGTGT 1656
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 1657 AAACAAGATAATTTCTTGGGAACAGCGCGCGCAGTAAAGTATAGCTGGTACGAAAAATGAT 1716
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1717 TATGTTACAGTGTCAATTTGGGTATATACGAGCGCTATTTTACTAAAGATGGTGAAGT 1776
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrGlnAsn 372
Db 1777 CTTGGTGGAAATGTTTTCTTTTGAACAACTACGATAACTCTAAAGTGATACCTCTAC 1836

```
QY 373 LeuGluLysArgAlaPheSerGlyGlyLe-----TrpTyrValArgAspArgAla 389
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1837 TATAAGCGTACGACTTACGGAAGTAATCTTACTTTCAGTTCCCTGTAATGAAATAAC 1896
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1897 TCCTATTATGTAAGATTAGGTACATCTATATAAATAATTAGTAACCTTGTCTGAGAATAT 1956
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 403 -----GlyArgLysIleProGlySerAsp 410
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1957 AACCGTAATTTATATATTCATCAATGAATTAAGGTAAATGGCATTAACAATAATGAC 2016
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2017 TTTCAT-----TTTCTTTTGGTTCGGAACATAACAGCCTT 2052
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2053 AATAGAGGCTATTTCCTCACTAAAGG-----GTTAAACCAAGCTCTTGTGTGACGAGTT 2106
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2107 ACTATTCCAGGTTCTGATAACAATACTACAACTAAGTCAGATGTACAGGCTTTCTAC 2166
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2167 CCATTAGACAGAGATCACCTCTGGGTGTCTATCTGCAAAAGCATCTGCAGGATATGCAAT 2226
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2227 GGTGTTTGGAAACAACGCGTTACCGTCTCTATCAAACTTATACAGCGGTGCGCATCGGTTCA 2286
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2287 TTACGTGGTTTTCGTTATGATGATTT-----GGACCTAACGCATTTATGCCGAA 2337
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 526 -----GlySerValLeuProGlu 531
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2338 TATGGTAATGGTAGGTGACTGTTTAAAGAAATAAGTTCTCATCTGATGTGTGT 2397
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 532 ArgAlaLeuLeuValGlySerLeuGlyThrGlnLeuPro----- 544
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2398 AATGCAATCGCTACAGTACGCGAGGTAATTCGTGCCAATCCATTTGTGACGATAG 2457
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2458 AGCCAAATACGGTCCGAACCTCCTTATTTGTGATCGCGCAAGTGTTTGGAATACTAAA 2517
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2518 TGGAAATCAGATAAAATGGATTAGAGAGCGATGATTAAAAAGATTGCTGATTATGCG 2577
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2578 AATCAAGCGGTATTTCGCGCCTCTACAGGTGTCGGATTCCCAATGGCAATCTCCTATTGGG 2637
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2638 CCATTGGTATTCTTATGCCAAACCAATTAATAAATAATGAAATGATGATGTCGAA--- 2694
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 2695 CAGTCCCAATTTAGTATTGGAGGTTCTTTTC 2724
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
```

RESULT 11

```
US-09-135-166-5
; Sequence 5, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
```

```
; APPLICANT: STA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-09-135-166-5
```

Alignment Scores:

```
Pred. No.: 1.67e-12 Length: 2950
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24
```

US-09-857-669-2 (1-609) x US-09-135-166-5 (1-2950)

```
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 883 GAATCTGTTAGTAGCAGTACATTAACAAGACAAATGCAATTAACAACCTGATTCTTGTGG 942
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAsnValLys 100
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 943 AATATTGCGGAATAATTTGAAGGTGCGCAATTCGAGAAAGATTTCGACGATCATTCGT 1002
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1003 GATTATTATTAAATAATGGTATGCCAAAGCAAAATTAATAAAGCGATGTTTCAGCTA 1062
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1063 AATGATGAAAAACAAAGATTAAATGTAACCATTCATGTAATGAAGGTTTACAGTATGAC 1122
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db 1123 CTTTCGTAGTGCAGCATTATAGTAACTCTGGGAGGTATGTCGCCGAGCTTGAACCTTTA 1182
      :::::  ::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
```


[illegible]

Db 2518 TGAATACAGATAAAATGGATTAGAGCGGATGCTATTAAAGATTGCCCTGATTATGCC 2577
 QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
 Db 2578 AAATCAAGCGTATTTCGCCCTCTACAGGTGTCGATTCCAATGGCAATCTCTATTGGG 2637
 QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
 Db 2638 CCATTGGTATTCTTATGCGCAACCAATTAATAATATGAAATGATGATGTCGAA--- 2694
 QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
 Db 2695 CAGTTCCAATTAGTATGGAGGTTCTTTC 2724

RESULT 13

US-08-433-522A-3
 ; Sequence 3, Application US/08433522A
 ; Patent No. 6013514
 ; GENERAL INFORMATION:
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: THOMAS, Wayne
 ; APPLICANT: YANG, Yan Ping
 ; APPLICANT: LOOSMORE, Sheena
 ; APPLICANT: SIA, Dwo Yuan Charles
 ; APPLICANT: KLEIN, Michel
 ; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
 ; NUMBER OF SEQUENCES: 55
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SIm & McBurney
 ; STREET: 6TH Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/433,522A
 ; FILING DATE: 12-SEP-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEWART, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2984 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 374..2764

US-08-433-522A-3

Alignment Scores:
 Pred. No.: 1-7e-12 Length: 2984
 Score: 209.50 Matches: 130
 Percent Similarity: 35.85% Conservative: 103
 Best Local Similarity: 20.00% Mismatches: 280
 Query Match: 6.60% Indels: 137
 DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-08-433-522A-3 (1-2984)

QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrClnGlnGlu 80

Db 923 GAATCTGTAGTACAGCATACATTCAAGAACAAATGGAATTACAACTGATTCTTGGTGG 982
 QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
 Db 983 AAATATGGGGAATAAATTTGAAGGTGGCCAATTTCAGAAAGATTTCGACAGTCAATCGT 1042
 QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
 Db 1043 GATTATTATTAAATAATGCTATGCCAAGCACAANTTACTAAACCGATGTTTCAGCTA 1102
 QY 117 -----GluLysAspGlyAla-----TyrThrValHisIleThrProGlyProArgThrLys 133
 Db 1103 AATGATGAAAAACAAAAGTTAATGTAACCAATTGATGTAATCAAGATTTCACAGTATGAC 1162
 QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
 Db 1163 CTTGCTAGTCACGCCATTATAGGTAATCTGGGAGGTATGCTGCCGAGCTTGAACCTTTA 1222
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluGluAsnTrpGlnGlnPro 165
 Db 1223 CTTTCAGCATTTACATTTAAATGATACTTTCCGCCGTAGT----- 1261
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
 Db 1262 -----GATATTGCAGATGTAGAAAATGCAATTAAGCAAAACTTGGG--- 1303
 QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 Db 1304 -----GAACGCGGTTACGGTAGCGCAACGTAATTCAGTA-----CCT 1342
 QY 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
 Db 1343 GATTTGTGATGCAATAAATACATTAGCGATTAACCTTGTGTGTGATGCTGGACGAGT 1402
 QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
 Db 1403 TTAACGTGTTCCCAACTTCGCTTTGAAGGAATAATACCGTTTCGCTGATAGCACTTAGT 1462
 QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
 Db 1463 CAGGAAATGCCCAACAAAGAGGAACTTGGTATAATTCACATTTAGTTAGTTAGGAAAA 1522
 QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
 Db 1523 ATTCGCTTAGATCGTACAGGTTTCTTCGAA-----ACAGTCGAAAAACCGAATTGATCCT 1576
 QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
 Db 1577 ATCAATGGTAGTAATGATGAAGTGGATGCTGATATATAAGTCAAGAACCTAACACGGGT 1636
 QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
 Db 1637 AGTATCAACTTTGGTATTGGTTACGGTACACAGATGCTATTAGTTATCAAGCAAGTGT 1696
 QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
 Db 1697 AACACAGATAATTTCTTGGGACAGGGCGCGAGTAGTATAGCTGCTGAGAAAAATGAT 1756
 QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
 Db 1757 TATGTCAGAGTGTCAATTTGGGTTATATACCGAGCCCTATTTTACTAAAGATGGTAAAGT 1816
 QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
 Db 1817 CTGTGTGGAATGTTTTCTTTGAAAACTACGATACTCTAAAAGTGATACATCTCTAAC 1876
 QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
 Db 1877 TATAAGCGTAGCACTTACGGAAGTAATGTTACTTACTTGGTTTCCCTGTAATAATAAAC 1936
 QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402

```
Db 1937 TCCTATTATAGGTAGGTCATACCTATATAAATAAGTAACTTTCTCTAGAAATAT 1996
QY 403 -----GlyArgLysIleProGlySerAsp 410
Db 1997 AACCGTAATTTATATTCAATCAATGAATTTAAAGGTAATGGCAATTAACAAACATGAC 2056
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2057 TTTGAT-----TTTTCTTTTGGTGGAACTATAACAGCCTT 2092
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2093 AATAGAGCTATTTCCCAACTAAAGG-----GTTAAAGCAAGTCTTGCTGGACGAGTT 2146
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2147 ACTATTCCAGCTTCTGTAGCAAACTACTACAACTAAGTGCAGATGTACAGGGTTCTTAC 2206
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 2207 CCATTAGACAGAGATCCCTCTGGTTGTATCTGCAAAAGCATCTGCAGGATATGCAAA 2266
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 2267 GCTTTTGGAAACAAGCGTTTACCGCTTCTATCAAACTTATACAGCGGGTGGCATCGTTCA 2326
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2327 TTACGTGTTTTGCTTATGGTAGTATT-----GGACCTAACGCAATTTATCCCGAA 2377
QY 526 -----GlySerValLeuProGlu 531
Db 2378 TATGCTAATGCTAGTGTGCTACTGTTACTTTTAAAGACATAAGTCTGATGTGATCGTGT 2437
QY 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2438 AATGCAATCGGTACAGTACGCGAGAGTTAATTTGTGCCAACTCCATTTGTGAGCGATAAG 2497
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2498 AGCCAAATACGGTCCGAACCTCTTATTTGTTGATGCGCGCAAGTGTGGAATACTAAA 2557
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2558 TGGAAATCAGATAAAATGGATTAGACAGCGGATGATTAAAGAGATTGCGCTGATTATGGC 2617
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2618 AAATCAAGCCGTTATTCGCGCTCTACAGGTGTCGGATTCCAATGGCAATCTCTATTGGG 2677
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2678 CCTTGGTATTCTCTTATGCGCAACCACTAAATAAATAATGAAATGATGATGTCGAA--- 2734
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2735 CAGTTCCCAATTTAGTATTGGAGGTTCTTTC 2764

RESULT 14
US-09-135-166-3
; Sequence 3, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
```

```
; STREET: 6TH Floor, 330 University Avenue
; City: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..2764
; US-09-135-166-3

Alignment Scores:
Pred. No.: 1,7e-12 Length: 2984
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137
DB: 3 Gaps: 24

US-09-857-669-2 (1-609) x US-09-135-166-3 (1-2984)
QY 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlu 80
Db 923 GAATCTGTTAGTACGACGATTCACAGACAATGGAATTACAACCTGATTCCTGGTGG 982
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 983 AAATTTATGGGAAATAAATTTGAAGGTGCGCAATTTCGAGAAAGATTGCGATCAATTCGT 1042
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr----- 116
Db 1043 GATTATTATTAAATAATGCTATGCCAAAGCACAAATTAATAAAGCGATGTTTCAGCTA 1102
QY 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
Db 1103 AATGATGAAAAACAAAGGTTAATGTAACCATTCATGTAATGAAAGGTTTACAGTATGAC 1162
QY 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
Db 1163 CTTCGTAGTCACCCATTATAGGTAATCTGGAGGTTATGCTGCCGAGGTTTGAACCTTTA 1222
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 1223 CTTTCAGCATTCATTAAATGATGACTTTCCGCCGCTAGT----- 1261
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
```

Db 1262 -----GATATTGCAGATGTAGAAAATGCAATTAACCAAAACTTGA--- 1303
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaValAsnPro 205
Db 1304 -----GAACGGGTTCGGPAGCGCAACGGTAATTCAGTA-----CCT 1342
QY 206 Asp-----ThrAlaThrValAspLeuAsnValValAspSerGlyArgPro 221
Db 1343 GATTTTGGATGCAATAAATAGCGATACCCCTGCTGTCGTCGACGAGCT 1402
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
Db 1403 TTAACGTGTCGCCAACTTCGCTTTGAAGGAATAACCGTTCTGCTGATAGCACTTTAGCT 1462
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
Db 1463 CAGGAATGCCCAACAGAGAACTTGGTATATTAATTCACATTAAGTTAGTAGGAAAA 1522
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1523 ATTCGCTTAGATCGTACAGGTTCCTCGAA-----ACAGTCGAAACCGAAATTCATGCT 1576
QY 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
Db 1577 ATCAATGGTAGTATGATGCAAGTCGCTATATTAAGTCAAGAACGTAACACGGGT 1636
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyLysIleAlaTyr 319
Db 1637 AGTATCAACTTTGGTATTGGTTACGAGACAGAGGTGATTATTAACAAGCAAGTGT 1696
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLys 339
Db 1697 AAACAAGATAATTTCTGGGAACAGGGCGGCGAGTAAGTATAGCTGCTACGAAATGAT 1756
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1757 TATGTCAGAGGTGCAATTTGGGTTTATACCGAGCCCTATTTTACTAAGATGGGTAGAT 1816
QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1817 CTTGGTGGAAATGTTTCTTTGAAACATACGATAACCTCAAAAGTGCATACCTCTAAC 1876
QY 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389
Db 1877 TATAAGCTAGACACTTACGGAAGTAATGTTACTTATGTTTCCCTTAATGAAATAAC 1936
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1937 TCCTATTATGTAGGATTAGGTCATACCTATAATAAATAGTAACCTTCTCTAGATAT 1996
QY 403 -----GlyArgLysIleProGlySerAsp 410
Db 1997 AACCGTAATTTATATTTCAATCAATCAATGAAATTTAAAGGTAATGGCAATTAACAAATGAC 2056
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2057 TTTGAT-----TTTTCTTTTGGTTGGAACTATAACAGCCTT 2092
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2093 AATAGAGGCTATTTCCCAACATAAAGG-----GTTAAAGCAAGTCTTGGTGACGAGTT 2146
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2147 ACTATTCCAGGTTCTGATAACAAATACTACAACCTAAGTGCAGATGTCACGGGTTTCTAC 2206
QY 471 ThrProGluAsnLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
Db 2207 CCATTAGACAGAGATCAGCTCTGGGTGTATCTGCAAAAGCATCTCAGGATATGCAAT 2266
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 2267 GGTTTTGGAAACAACGGTTTACCCTTCTATCAAACTTATACACCGGTGGCATCGTTCA 2326

QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2327 TTACGTGGTTTGGCTTATGGTAGTATT-----GGACCTAACCAATTTATGCCGAA 2377
QY 526 -----GlySerValLeuProGlu 531
Db 2378 TATGCTAATGGTAGGTGCTGCTTTTAAAGAAGATAGTCTCATGCTGATTGTTGCT 2437
QY 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2438 AATGCAATCGCTACAGCTAGCGAGAGTAAATTTGTGCAACTCCATTTCTGAGCGGATAAG 2497
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 2498 ACCCAAAATACGGTCCGAACCTCTTATTTGTTGATGCGCGCAAGTCTTTGGAATACTAAA 2557
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2558 TGGAAATCAGATAAAATTCGATTAGAGAGCGATGTATTAATAAGATTGCTGATTATGCG 2617
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAla 584
Db 2618 AATCAACCGGTATTCGGCGCCTCTACAGGTGTCGGATTCGAATGGCAATCTCTATTGGG 2677
QY 585 PropheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2678 CCATTTGGTATTCTTATTCGCCAACCAATTAATAAATATGAAATGATGATGTCGAA--- 2734
QY 600 ArgTyrPheHisLeuSerLeuGlyThrArgPhe 609
Db 2735 CAGTCCCAATTAGTATTGGAGGTCTTTC 2764

RESULT 15

US-08-942-046-3
; Sequence 3, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163

[illegible]

RESULT 17

```

US-09-135-166-55
; Sequence 55, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0.
; Version #1.25

```


QY 242 GlyLeuAlaAaGpHeGlnProGlyThrProTyrAspLeuAspLeuLeuLeuAspPheGln 261
Db 1478 CAGGAATGCGCAACAAGAAAGGAGCTGGTATATTAATTCACATAGTTGAGTTAGGAATA 1537
QY 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
Db 1538 ATTGCGTTAGATGTCACAGGTTTCTCGAA-----ACAGTCGAAACCGAATTGATCCT 1591
QY 282 LeuGlnGly-----AspArgValProValIysValSerValThrGluValIysArgHis 299
Db 1592 ATCAATGGGTAGTATGATGAAGTGGATGTGTATATAAGTCAAGAAGCTAACCGGGT 1651
QY 300 LysLeuGluThrGlyLeuArgLeuAspSerGluTyrGlyLeuGlyGlyIysLeuAlaTyr 319
Db 1652 AGTATCAACTTGGTATGGTTACCGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGT 1711
QY 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValThrAspMetAspLys 339
Db 1712 AACAAAGATATTTCTGGGAACAGGGCGCGCAGTAAATATAGCTGGTACGAAATATGAT 1771
QY 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
Db 1772 TATGGTACAGGTGCAATTTGGGTATACCGAGCCCTATTTTACTAAAGATGGTGTAACT 1831
QY 354 TyrArgGlyAsnTyrThrThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
Db 1832 CTGTGGGAATGTTTCTTGAACAACTACATAACTCTAAAGTATGATCCTCTCTAAC 1891
QY 373 LeuGluLysArgAlaPheSerGlyGlyLe-----TrpTyrValArgAspArgAla 389
Db 1892 TATAAGCGTACGACTTACGGAAGTAAGTACTTACTTGTAGTTTCTCTGTAATGAAATAAC 1951
QY 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
Db 1952 TCCTATTATGTAGGATTAGGTACACTATATAAATAATAGTAACTTCTCTCTAGATAT 2011
QY 403 -----GlyArgLysIleProGlySerAsp 410
Db 2012 AACCGTAATTTATATATTCATCAATCAATGAAATTTAAAGGTATGGCATTAACAAATGAC 2071
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 2072 TTTGAT-----TTTTCTTTTGGTGGAACTATACAGCCCT 2107
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 2108 AATAGAGGCTATTTCCCACTAAAGG-----GTTAAAGCAAGTCTGTCGACGAGTT 2161
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 2162 ACTATTCAGGTTCTGATAACAATACTACAACTAGTACAGATGTACAGGTTTCTTAC 2221
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrVal 490
Db 2222 CCATTACAGAGATCAGCTCTGGGTGTATCTGCAAAAGCATCTGAGGATATCAAT 2281
QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
Db 2282 GGTTTTGGAAACAACGGTTTACCGTTCTATCAAACTTATACAGCGGGTGCATCGGTTCA 2341
QY 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
Db 2342 TTACGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2392
QY 526 -----GlySerValLeuProGlu 531
Db 2393 TATGGTAAATGCTAGTCTAGTCTATTTTAAAGAGATAAGTCTGATGGTGGTGGT 2452
QY 532 ArgAlaLeuValGlySerLeuGluTyrGlnLeuPro----- 544
Db 2453 AATGCAATCGCTACAGCTAGCGCAGATTAATTTGTCGCAACTCCATCTGTGAGCGATAAG 2512
QY 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557

Db 2513 ACCCAAAATACGGTCCGAACCTCTCTTATTTGTTGATGCGCGCAAGTCTTTGGAACTACTAAA 2572
QY 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
Db 2573 TGGAAATCAGATAAAATGGATAGAGAGCGGATCTATTAAAAAGATTGCTGCTATTATGGC 2632
QY 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
Db 2633 AAATCAACCGGTATTCGGCCCTCTACAGGTGTGGATTCATGGCAATCTCTATTTGGG 2692
QY 585 ProPheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
Db 2693 CCATTTGGTATTTCTCTTATGCGCAACCAATTAATAAATATGAAATGATGATGTCGAA--- 2749
QY 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
Db 2750 CAGTTCCAATTAGTATTTGGAGGTTCTTTC 2779

RESULT 18
US-08-942-046-55
Sequence 55, Application US/08942046
Patent No. 6264954
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS:Jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2987 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-942-046-55

Alignment Scores:
Pred. No.: 1.7e-12 Length: 2987
Score: 209.50 Matches: 130
Percent Similarity: 35.85% Conservative: 103
Best Local Similarity: 20.00% Mismatches: 280
Query Match: 6.60% Indels: 137

DB: 4 Gaps: 24
US-09-857-669-2 (1-609) x US-08-942-046-55 (1-2987)
Qy 61 AspSerGluIleLysAspMetValGluGluHisLeuProLeuThrGlnGlnGlu 80
:::||||| :::::||||| :::::|||||
Db 938 GAATCTGTTAGTAGCAGTACATACAGAAACAAATGGAATTAACAACCTGATTTCTGGTG 997
Qy 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLys 100
:::||||| :::::||||| :::::|||||
Db 998 AATTATGGGGAATAAATTTCAAGCTGCCAATTCAGAAAGATTTCCAGTCAATTCGT 1057
Qy 101 ThrMetLeuArgSerLysGlyThrPheSerSerLysValSerLeuThr----- 116
:::||||| :::::||||| :::::|||||
Db 1058 GATTATTATTAAATGCTATGCCAAAGCACAATAATTAACAAACGGATGTTTCAGCTA 1117
Qy 117 -----GluLysAspGlyAla---TyrThrValHisIleThrProGlyProArgThrLys 133
||||| :::::||||| :::::|||||
Db 1118 AATGATGAATAAACAAGATTAACTGATACCAATGATGTAATGAGGTTTACAGTATGAC 1177
Qy 134 -----IleAlaAsnValGly-----ValAlaIleLeuGlyAspIle 145
||||| :::::||||| :::::|||||
Db 1178 CTTCCGTAGTCACGCAATATAGGTATATCTGGGAGGTATGCTCGCAGCTTGAACCTTTA 1237
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrThrArgAsnAlaLeuGluAsnTrpGlnPro 165
||||| :::::||||| :::::|||||
Db 1238 CTTTCACATTACATTAATGATTAATGATCTTCGCCGCTAGT----- 1276
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
||||| :::::||||| :::::|||||
Db 1277 -----GATATTGCAGATGAGAAATGCAATTAAGCAAAACTTGGAA--- 1318
Qy 186 ValThrArgLysGlyThrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
:::||||| :::::||||| :::::|||||
Db 1319 -----GAGCGGTTACGTAGCCACGCGTAATTCAGTA-----CCT 1357
Qy 206 Asp-----ThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
GATTTTGGTATGATGCAATAAATACATTAACGCTTGTGTTGATGCTGCGACGCT 1417
Qy 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgYrProGluGlnIleValSer 241
:::||||| :::::||||| :::::|||||
Db 1418 TTAACGTTCGCCAATTCGCTTTCGAGGAATAACCGTTCTGCTGATACCACTTTACGT 1477
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
:::||||| :::::||||| :::::|||||
Db 1478 CAGGAATGCGCAACAGAGAACTGGTATATATTCACAAATAGTTGAGTTAGAGAAA 1537
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
:::||||| :::::||||| :::::|||||
Db 1538 ATTCGCTTAGTCGATACAGTTCTTCGAA-----ACAGTCGAAACCCGAATTGATCCT 1591
Qy 282 LeuGlnGly-----AspArgValProValLysValSerValThrGluValLysArgHis 299
:::||||| :::::||||| :::::|||||
Db 1592 ATCAATGGTAGTAATGATGAAGTGTGCTATATATAAGTCAAGAAACGTAACACGGGT 1651
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
:::||||| :::::||||| :::::|||||
Db 1652 AGTATCAACTTTGGTATTGGTTACGGTACAGAGTGGTATTAGTTATCAAGCAAGTGT 1711
Qy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
:::||||| :::::||||| :::::|||||
Db 1712 AACACAGATAATTTCTTGGGAACAGGGCGGCAAGTAAATATAGTGTGCTAGCAAAATGAT 1771
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg-----Asn 353
:::||||| :::::||||| :::::|||||
Db 1772 TATGGTACGAGTGTCAATTTGGTTATACCGAGCCCTATTTTACTAAAGATGCTGTAAGT 1831
Qy 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSer---ThrThrGlnAsn 372
:::||||| :::::||||| :::::|||||
Db 1832 CTTGGTGGAAATGTTTCTTCTTGGAAACATACGATAACTTAAAGATGATACCTCTAAC 1891
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIle-----TrpTyrValArgAspArgAla 389

Db 1892 TATAACGGTACGACTTACGGAAGTAATGTTACTTTAGTTTCCCTGTAATAAGAAATAAC 1951
Qy 390 GlyIleAspAlaArgLeuGly-----AlaGluPheLeuAlaGlu--- 402
:::||||| :::::||||| :::::|||||
Db 1952 TCCTATTATCTAGGATTAGGTATACCTATAATAAATAGTAACCTTTCCTAGAAATAT 2011
Qy 403 -----GlyArgLysIleProGlySerAsp 410
:::||||| :::::||||| :::::|||||
Db 2012 AACCGTAATTTATATATTCATCAATCAATGAATTTAAAGTAATGCGCATTAACAACATGAC 2071
Qy 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
:::||||| :::::||||| :::::|||||
Db 2072 TTTGAT-----TTTTCTTTTGGTTGGAACTATAACAGCCTT 2107
Qy 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
:::||||| :::::||||| :::::|||||
Db 2108 AATAGAGGCTATTTCCCACTAAAGCG-----GTTAAACCAAGCTCTTGGTGACAGTT 2161
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
:::||||| :::::||||| :::::|||||
Db 2162 ACTATTCCAGGTTCTGATAACAAATACTACAACCTAAGTCCAGATGTACAGGGTTTCTAC 2221
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal 490
:::||||| :::::||||| :::::|||||
Db 2222 CCATTAGACAGACATCACCTCTGGGTTGTATCTGCANAGCATCTCGAGATATGCAAT 2281
Qy 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
:::||||| :::::||||| :::::|||||
Db 2282 GCTTTTGGAAACAAGCGTTTACGCTTCTATCAAACTATATACAGCGGTGCGATCTCA 2341
Qy 511 ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn----- 525
:::||||| :::::||||| :::::|||||
Db 2342 TTACGTGGTTTCTTATGCTAGTATT-----GGACCTAACCCAATTTATGCCGAA 2392
Qy 526 -----GlySerValLeuProGlu 531
:::||||| :::::||||| :::::|||||
Db 2393 TATGGTAATGGTAGTGGTACTGGTACTTTTAAAGAGATAAGTTCTGATGTGATGGTGT 2452
Qy 532 ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro----- 544
:::||||| :::::||||| :::::|||||
Db 2453 AATGCAATCGCTACAGTACGACGACAGATTAAATTGTCCAACTCCATTTGTGAGCGATAAG 2512
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
:::||||| :::::||||| :::::|||||
Db 2513 AGCCAAATACGGTCGGAACCTCTTATTTGTTGATGCGCAAGTGTGTTGGAATACTAA 2572
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMet----- 567
:::||||| :::::||||| :::::|||||
Db 2573 TCGAAATCAGATAAAATGGAATTACAGACGCGATGTATTAAAAAGATTGCGCTGATTGCG 2632
Qy 568 -----LysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAla 584
:::||||| :::::||||| :::::|||||
Db 2633 AAATCAAGCGTATTGCGCGCTCTACAGGTGTGCGATTCCCAATGCGCAATCTCTATTGGG 2692
Qy 585 PhePheSerPheAspIleAla-----TyrGlyHisSerAspLysLysIle 599
:::||||| :::::||||| :::::|||||
Db 2693 CCATTGCTATTCCTTATGCCAACCAATTAATAAATATGAATATGATGTGCGAA--- 2749
Qy 600 ArgTrpHisIleSerLeuGlyThrArgPhe 609
:::||||| :::::||||| :::::|||||
Db 2750 CAGTTCCCAATTTAGTATTGGAGGTTCTTTTC 2779

RESULT 19

US-09-346-408-5
: Sequence 5, Application US/09346408B
: Patent No. 6338966
: GENERAL INFORMATION:
: APPLICANT: Allen, Steve
: APPLICANT: Anderson, Shawn
: APPLICANT: Falco, Carl
: APPLICANT: Rafalski, Antoni
: TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins


```

: TYPE: DNA
: ORGANISM: Momordica charantia
US-09-346-408-3

Alignment Scores:
Pred. No.: 0.00381 Length: 2280
Score: 121.50 Matches: 98
Percent Similarity: 35.70% Conservative: 63
Best Local Similarity: 21.73% Mismatches: 177
Query Match: 3.83% Indels: 113
DB: 4 Gaps: 24

US-09-857-669-2 (1-609) x US-09-346-408-3 (1-2280)
Qy 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
Dy 317 CCCAGAAATCCAGATCTCCAGCATCGATATCCAGTGGGT---CATGTGCTCAGCGAAGC 373
Qy 150 AsnLeuAlaCluTyrrArgAsnAlaLeuGluAsnTrpGlnGlnProVal-----166
Dy 374 -----TGGCCAGCCCTCTCACCGGATTC 397
Qy 167 -----GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Dy 398 ATCAGAGAAATCCGAGTTCCTCCAAACGCTTCATTTCAATTCCTCCAGGCTCCCGGATGG 457
Qy 185 AlaValThrArgLysGlyTyrrProLeuAla-----LysLeuGlyAsnThrArg 200
Dy 458 TCTGTGGCGAATATGTCAGTGGCCCATTTCTTCGCCCATTTGACGATGCCCAAAAGCACCGC 517
Qy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Dy 518 ATCGCCGACTCCACTTCTGCTGCTCTCTTTGAGGCCAAC-----AACAAAT 562
Qy 221 ProIleAla---PheLysAspPheGluIleThrGlnArgTyrrProGluGln---238
Dy 563 CCCATGCCCATTTCTCAAGATATTGAGATC-----TACAAGCACCTTGAAGAAGAA 613
Qy 239 ---IleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrr-----252
Dy 614 AGCATAGCAAGAACCTGGGGCCACCCTGCCCGCAGACTCCCTTACGTTGATCAAGCTATA 673
Qy 253 -----AspLeuAspLeuLeu-----LeuAspPhe 260
Dy 674 ACCAATGCTGGTAATTGCGTGATTTGGGGGTGATTTGGAGGTATAGAACCAATCAAGTAC 733
Qy 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrrSerGlyAlaSerValGlnAlaAspPheAsp 280
Dy 734 CATGATGGTCTGGACCGT---TTCCGACAGTCACCTGCAGAACTACGGGAGGAATTTACC 790
Qy 281 ArgLeuGlnGlyAspArgVal-----ProVal-----289
Dy 791 CGGGGAATGCGATGCGATTTTCATTCCTCAACTTCGCAATCCAGTCCATACCGGCAT 850
Qy 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg-----Leu 307
Dy 851 GCTTTACTAATGACTACACCGCCGCTCGGCTGCTCGATATGGGTACAAAGACCCCAT 910
Qy 308 AspSerGluTyrrGlyLeuGlyLysIleAlaTyrrAspTyrrAsnLeuPheAsnLys 327
Dy 911 TTGTGCTTCATCTCTGGTGGC-----TACACCAAG 943
Qy 328 GlyTyrrIleGlySerValTrpAspMetAspLysTyrrGluThrLeuAlaIleGly 347
Dy 944 GCAGATGATGTTCCACTATTGTTGGGGAATGAAGCAACATGAGAGAGTCTTGAGGATGGT 1003
Qy 348 IleSerGlnProArgAsnTyrrArgGlyAsnTyrrTrpThrSerAsnValSerTrpAsnArg 367
Dy 1004 GTTCTGTGATCCAGACACAGGTGTCTCCATATTTCCATATTTCCATATTCGACTATGAGGG 1063
Qy 368 SerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrrValArgAsp 387
Dy 1064 CCAACAGAGGTGCAG-----TGGCATGCAAAAGCG 1093

```

```

Qy 388 ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIlePro 407
Dy 1094 AGA-----ATCAATGCA-----GGGGCTAACTTTTACATTTGTTGGCCGTGAC-----1135
Qy 408 GlySerAspIleAspLeuGlyAsnSerHisAlaThr-----MetLeuThrAla 423
Dy 1136 -----CCTGCTGGAATGGGTCTACTCTAGAGAAAGGGATCTCTATGATGCT 1183
Qy 424 SerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrrLeuAsp 443
Dy 1184 GACCATGGGAAGAAAGTACTGAGCATGGCAGCTGGAGCGGTAAACACTCTTCTCT 1243
Qy 444 GlyLysIleGlyThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSer 463
Dy 1244 TTCAGAGTT-----GCTGCTTATGATAAACTCAG 1273
Qy 464 AlaArgAlaGlyTyrrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg 483
Dy 1274 GGCAAAATGGCC---TTTTCGATCCCTCGAGACTCAGGATTTCTTTTC---ATATCC 1327
Qy 484 GlyGlnAlaGlyTyrrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPhe 503
Dy 1328 GGCACCAAGATGAGAACACTAGCAAGAAAGAGAAATCCACCAGAGAGGTTTATGTGC 1387
Qy 504 ArgSerGlyGlyAlaSerSerValArgGlyTyrrGluLeuAspSerIleGlyLeuAlaGly 523
Dy 1388 CCAGGGGGATGGAGGTGTTGTTGAATATTATGAC-----AGTTTGGTACCT 1435
Qy 524 ProAsnGlySerValLeuProGluArgAlaLeu 534
Dy 1436 GCAAGCAATGACAGACTTCTCGAAGCTGTCTTA 1468
RESULT 21
US-09-346-408-7
; Sequence 7, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Anderson, Shawn
; APPLICANT: Falco, Carl
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
; FILE REFERENCE: BB-1167-A
; CURRENT APPLICATION NUMBER: US/09/346.408B
; EARLIER FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 60/092,833
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1890
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-346-408-7

Alignment Scores:
Pred. No.: 0.00866 Length: 1890
Score: 117.00 Matches: 110
Percent Similarity: 32.53% Conservative: 52
Best Local Similarity: 22.09% Mismatches: 200
Query Match: 3.69% Indels: 136
DB: 4 Gaps: 26

US-09-857-669-2 (1-609) x US-09-346-408-7 (1-1890)
Qy 104 ArgSerLysGlyTyrrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrrThr 123
Dy 112 CGCGCGCCCGCATGTCGGCCATCCGACCTCGCTCATCGACGAGCGGCGCGCTC 171
Qy 124 ValHisIleThr-----ProGly-----129
Dy 172 GTCGACCTCTGCGCGCGCGGAGCGCGCGGCTGCGCGGCGGAGCGGCGGCGGCGCTC 231

```

```
Oy 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuLeuGlyAspIleLeuSerAspGly 149
Db 232 CCGCGGTGGCGGTCCGCGCGCGGTGGACGGT----- 261
Oy 150 AsnLeuAlaGluTyrArgAsnAlaLeuLeuGluAsnTrpGlnProValGlySer--- 168
Db 262 -----GAGTGGCGCACGTGCTCGCGAGGGCTGGCGCTCCCGCGTCCGCGGCTC 312
Oy 169 -----AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 313 ATCGGGAGACAGATACCTCCAGTCCTCCACTTCACTCCCTCCGCTCCCTCCGCG 372
Oy 185 AlaValThrArgLysGlyTyrProLeuAla-----LysLeuGlyAsnThrArg 200
Db 373 GGCCTCCCAACATGCTGCTCCCATCGTCTCGCGCTCGACGACGCGCGCGCAAGACGC 432
Oy 201 AlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
Db 433 GTCGCGCGCGCGCGCGACGTCGCG----- 456
Oy 221 ProIleAlaPheGlyAspPheGluIleThrGlnArg-----TyrPro 236
Db 457 ---CTCGCGCGCGCGCGAGCGAGCTCCTCGCGCTCCTCGCGAGTTCGAAATATACCT 513
Oy 237 -----GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyr 252
Db 514 CACAATAAAGAAGAAGGATGCAAGAACATGGGGACACTGCCCTGGCTTACCTATAT 573
Oy 253 -----AspLeuAspLeuLeu 257
Db 574 GTCGATGAGCGGATACACCCAGCTGGAACCTGGCTGATTGTGTGTGATCTGGAGGTGTG 633
Oy 258 LeuAspPheGlnGlnAlaLeuGluGlnAsn-----GlyHisTyr-----SerGly 272
Db 634 -----CAKCCATTAAGTAAACGATGGCTTGACCATTCAGACGCTTCACCC 681
Oy 273 AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVal----- 287
Db 682 CAGCAACTTAGGACGAATTCGACAAACGCTGGGCTGTGTTATTTGCATTCCAGTTG 741
Oy 288 -----ProVal-----LysValSerValThrGluValLysArgHisLysLeu 301
Db 742 AGAAACCCAGTCCCAATGGCGATGCACACTGTTGATGATGACATGACATGAGCGCTCTTG 801
Oy 302 GluThrGlyIleArg-----LeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr 319
Db 802 GAAATGGGTTTCAAGATCCCATCTACTGCTACACCCCTTGGGTGGT----- 849
Oy 320 AspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 850 -----TTTACAAAGCTGATGATGCTCCGCTGCTGTAGATGGACAA 894
Oy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgLysAsnTyrTrp 359
Db 895 CACAGCAAGGTCTTAGAAGATGGAGTCTTGACCCCGAGACCACTATCGTCTATATTT 954
Oy 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
Db 953 CCCTCCCAATGCATATGCTGCTCCAGCAGAGTGCAG----- 993
Oy 380 GlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
Db 994 -----TGGCATGCAAGGACGA-----ATTACGCC-----GGTGTAAATTC 1032
Oy 400 LeuAlaGluArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
Db 1033 TACATAGTGGTCTGTGAT-----CCAGCTGGGATGGCGCATCCCGACA 1074
Oy 420 MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGly 439
Db 1075 GAG-----AAGAGAGATCTGTACAAC-----CCAGACCATGGG 1107
```

```
Oy 440 ---HisTyrLeuAspGlyLysIleGly---ThrThrLeuGlyThrPheLeuSerSerThr 457
Db 1108 AAGAGGTCTTAAGCATGCGCCCGGTTTGGAGAACTCAACATATTGCCCTTCAAGGTA 1167
Oy 458 AlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeu 477
Db 1168 GCAGCATATGATACGGTGGCCAAAGATGGCTTTCTTTGAACCTTCACGCAGTCAAGAT 1227
Oy 478 GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspVal 497
Db 1228 TTTCTGTGTC---ATCTCAGGAACCAAGATGCGCACTTTCGCCAAACTGGGAGAACCT 1284
Oy 498 ProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyr---GluLeu 516
Db 1285 CCTGATGTTTCATGTCCTGCTGGTGAAGGTCTTGTGTGACTACTACATAGCTTG 1344
Oy 517 AspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeu 534
Db 1345 CAAACTGAAGAGCTACCGCCCGCGCTGCTACTGTATGAGACAGCTGCTG 1398
RESULT 22
5223391-4/c
; Patent No. 5223391
; APPLICANT: COEN, DONALD M.; DIGARD, PAUL E.
; TITLE OF INVENTION: INHIBITORS OF HERPES SIMPLEX VIRUS
; REPLICATION
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/482,634
; FILING DATE: 21-FEB-1990
; SEQ ID NO:4:
; LENGTH: 1654
5223391-4
Alignment Scores:
Pred. No.: 0.0131 Length: 1654
Score: 114.50 Matches: 69
Percent Similarity: 35.60% Conservatives: 41
Best Local Similarity: 22.33% Mismatches: 109
Query Match: 3.61% Indels: 90
DB: Gaps: 10
US-09-857-669-2 (1-609) x 5223391-4 (1-1654)
Oy 121 AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAla 140
Db 1624 GCGCACCAGATCCACGCGCTTGATGACATCTTACCCCGTAGATGACGCGGATGCTACT 1565
Oy 141 IleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGlu 160
Db 1564 TTTCTTGGCGAT----- 1553
Oy 161 AsnTrpGlnGlnProValGly-SerAspPheAspGlnAspSer-----TrpGluAs 177
Db 1552 -----CAGCAGCAGCTTGTGAACGCTCTTTTCGCACTCGAGTTGATGGGCGCAGAAA 1499
Oy 177 nSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuG 197
Db 1498 CAGCGCGCGGAGATGGCTCGCCATC-----TTGTGCGCCCATGGG 1457
Oy 187 yAsnThrArgAlaAlaValAsnPro-AspThrAlaThrVal-AspLeuAsnValValVal 216
Db 1456 GGTGAGCGCGCGCGCTGAGCGCGCGGACAGCAAAATATGAGTCCGTGCTCCCGTA 1397
Oy 217 AspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPro 236
Db 1396 GATGATGCGCATGGAATAGGCGCGCGCGCATGTCGCGCGCTCGGGAAATCGCG 1337
Oy 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 256
Db 1336 CAGGAGTGTTCGAA-----GGCGCCCGCAGCGCGGTGACGTA 1298
Oy 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
```

```
Db 1297 CTCGCGGCTCGGAGCAGCATCTCGCGCGGATGGTCTCACCGTCGCGGCAACGTGCAG 1238
Qy 277 AlaaspPheaspArgLeuGlnGlyaspArgValproValLysValSerValThrGluVal 296
Db 1237 GCACGGCAGCAGCTCGGTCTGCACTCCCGTGAACCGGTACAC---CGAGTTACA----- 1187
Qy 297 LysArgHisLysLeuGluThrGlyLeuArgLeuAspSerGluTyrGlyLeuGlyLys 316
Db 1186 -----CACGACCTTGATGGCGGCTGCTGCTTGTCCAGGACGACGGCCTC----- 1142
Qy 317 IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
Db 1142 ----- 1142
Qy 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyLeuSerGlnProArgAsnTyrArgGly 356
Db 1141 -----CTCGGGCTGCTCTCGGGAATCCGGA----- 1115
Qy 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 1114 -----GCGGATCTGCTTTCGCATGGCGAGCCAGTCCCGCAGGAGGATGCTGAGCAG 1064
Qy 377 AlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyLeuAspAlaArgLeuGly 396
Db 1063 GCTCTCTCGCAGGCGCTC-----CTCGGGGCTGCTCTGCGG 1028
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 1027 AATCGCGAGCGGATCTGCTTTCGGATGGCGAGCCAGTCCCGCAGGAGGATGCTGAGGAG 968
Qy 414 GlyAsnSerHisAlaThrMet 420
Db 967 GCTCTCTCGCAGGTGAGCGT 947
```

```
RESULT 23
US-08-038-682-5
; Sequence 5, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-038-682-5
Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservativeness: 74
Best local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 1 Gaps: 33
US-09-857-669-2 (1-609) x US-08-038-682-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
Db 1529 ATATCGCTAAACCGGTGGT---GTGGAGACGTCGGGCGATGTTTATTTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACGCCAAAGAGTGGTTGTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGATATATGTATCTATTATTCAGCAAAACACA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GCGAGC-----AGCAATACTTCAGAAAGAGATGAATACACGGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTGCACGACCCCAAAACGAAACAAACAAACATTTAAACAACAACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATATAAAAAAGGTACC-----TTTGTAAACATCTACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCAATAGCTCCATTAATTTATCCAAATCGCAGCTTAACCTTTGGAGTCAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTGCGGCGGT-----GGCGGCGGT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTAAACAGATATTACCACCGGTGATCATACCAGAGGTGCRAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAAACAATTTACTCAGCGCGGTGGGTGTGATGTTTCATAAAATATCTCACTCGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTACATAAACATTTACAGCTAAACAGATATCGCCTTTGAGNAAGGAAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGCTCAAGGGACTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTTAGATTTTAATATGCTCTCTAAACGGCACTGCG 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGAGCTGCAA-----TTCAACCACTAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 317
```

```
Db 2170 AATAGCTCATCAAAATAAATTTGAAGGACCTTTAAATATTTTCAGGAAAGTGAACATC 2229
QY 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTTTTACCTAAATAAATAAGATGGATATGATATAATCAAGAGCGACCTTACTGG 2289
QY 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTTAACTCGAAAGTGATATGATAAATCAAGAGCGCCCTCACTATTACTCCAGA 2349
QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTCAGGCACACTTTACCCAGCCTTATAATTTAAACCGT----- 2397
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGlyLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTTCAACAAAGACACTACCTTTAATGTTGAACGAAATGCA----- 2442
QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCAACTTTGACATCAAGCGACCAATAGGATAAATAAGTAT 2487
QY 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTGAATTCAGCATCTTTAATGGAACACTTTTCAGTTTCGGGAGGGGAGTGTT 2547
QY 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTTC-----ACACTTCTCGCCTCATCC-----TCT 2574
QY 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AAGCTCAAAACCCCGGTAGTTATAAATCTCAATACITTAATCTTCAACAGGTCAC 2634
QY 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyThr 468
Db 2635 AGTTTAAGATTTAAACTTCAGGCTCAAGCTCAAAACACTGGCTCTCAAT-AGAGAAAGATT 2693
QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
Db 2694 AACTTTAAATGCCACCGGAGCAACATAACACTTTTCAAGTTGAAGGCCACCGATGGAAT 2753
QY 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 2754 GATTGTAAGCGATTGTAGCCAAATAAATAAATAACCTTTGAAGGAGGTAAGATG---AG 2810
QY 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGGAAGCGGTAAACA-----GAAATCGAAGGCAATGTTACTATCAATAA 2864
QY 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrGl 542
Db 2865 CAACGTACAGTC-----ACTTTATCGGTTCGGATTTCGACAAACATCA 2909
QY 542 nLeuProPheThr-----ArgThrLeuSerGl 551
Db 2910 AAAACCTTTTAACATATTAATAAAGATGTATCATTAATATAGCGCAACCTTACCGCTGGAGG 2969
QY 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
Db 2970 CAATATTGTCAATATAGCCGGAATCTTACCGTTGAAAGTAAGCTAATTTCAAA----- 3024
QY 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAAATTTCCAC 3041
QY 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAAATGTAGGC---GGCTGTTTGACAACAAGGAATTCAAATATTTCCTATGCCAA 3098
QY 606 ----GlyThrArgPhe 609
Db 3099 AGCAGGGGCTCGCTTT 3114
```

RESULT 24

```
US-08-302-832-5
; Sequence 5, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; OF NO. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-5
```

Alignment Scores:

```
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 136
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 1 Gaps: 33
```

US-09-857-669-2 (1-609) x US-08-302-832-5 (1-9171)

```
QY 47 valLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLleLysAsp 66
Db 1529 ATATCGCTAAACCGGTGTTT-GTGGAGACGTCGGGGCATGATGTTATTTCATCAAGAC 1587
QY 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGACCCCAAGAGTGGTGTGTAGAC----- 1623
QY 87 GlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGGATATGTTATCTATTAATGAGAAACAGCA 1656
QY 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
```

Db 1657 GGAGC-----AGCAATACCTTCAGAACGATGAATACACGGGATCCGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTGCAGACCCCAAAACGAAACAAAGAACACATTAACAACACAACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGGTATATAAAAAAGGTACC-----TTTGTAAACATCACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGTCAATAGCTCCATTAATTAATCCAAATGGCAGCTTAACCTCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyValAlaThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTCCGAGCGGT-----GGCGGGCTT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTAAACAGCATATTACCACCGGTGATGATACACAGAGGTGCAAA 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTACTACGGCGGTGGTTCATGCTTCAATAAAATATCTCACTCGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAAATTTACAGCTAAACAGATATCGCCTTTGAGAAAGGAAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGGACTATTACCTCAGGCAT----- 2091
Qy 260 PheGlnGlnAlaLeuGlnAlaAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAGGTTTTAGATTTAATAATCTCTCTAAACGGCACTGGC 2136
Qy 280 AspArgGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTCCACCCTAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AATACGCTATCACAAATAAATTTGAGGGGACTTTAAATATTTACAGGAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTTTTACCTAAAAATCAAAAGTGGATATCATATAATTAAGAGCGCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspPheAspLysTyr 340
Db 2290 AATTTAACTCGAAAGTGGATGATGATAAATTTCAAGACGCCCTCACTATTGACTCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTGCAGGCACACTTACCAGCCTTATTAATTTAAACGGT----- 2397
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTCACAAAGACATACCTCTTTAATGTTTGAACGAATGCA----- 2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCAACTTTTGACATCAAGCACCACCAATAGGATTAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTCAATTTAGCATCATCTTAATGGAACATTTTCACTTTCCGGAGGGGGAGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTT-----ACACTTCTCGCCTCATCC-----TCT 2574

Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AAGTCAACACCCCGGTGTAGTTATAAATTTCTAAATACTTTAATGTTTCAACAGGGTCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 2635 AGTTTAAGATTTTAAACTTCAGGCTCAACAAAACTGGCTTCTCAAT-AGAGNAAGATTT 2693
Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
Db 2694 AACTTTAAATGCCACGGAGGCAACATAACACTTTTTCGAAGTTTGAAGGCCGATGGAAT 2753
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGGTAAAGCATTTGTAGCCAAAAAACAATAACCTTTGAAGGAGGTAAAGTG---AG 2810
Qy 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGAAAGCGGTAAACA-----GAAATCGAAGGCAATGTTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrG 542
Db 2865 CRAAGCTAACGTC-----ACTCTTATCGGTTTCGATTTTGCACAAACNTCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerG 551
Db 2910 AAACCTTTTAACTATTAAAAAGATGTATCAATTAATAGCGGCAACCTTACCGCTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPhelysArgme 567
Db 2970 CAATATTGTCATATAGCGGAATCTTACCGTTGAAAGTAAAGCTAAATTTTCAAA----- 3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAATTTTCAC 3041
Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAAATGTAGC---GGCTTGTGTGACAAACAAAGGCAATTCAAATATTTCATTCGCCAA 3098
Qy 606 ----GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114
RESULT 25
US-08-530-198-5
; Sequence 5, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

; NAME: BERKSTRESSER, JERRY W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: JWB-1186
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9171 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-530-198-5

Alignment Scores:

Pred. No.: 0.23 Length: 9171
 Score: 114.00 Matches: 138
 Percent Similarity: 22.92% Conservative: 74
 Best Local Similarity: 21.36% Mismatches: 201
 Query Match: 3.59% Indels: 33
 DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-530-198-5 (1-9171)

QY 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLysasp 66
 DB 1529 ATATCGCTAAACCGGTGGTTT-GTGGAGACGTCGGGGCATGATTTATTCATCAAGAC 1587
 QY 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLysGlu 86
 DB 1588 AATGCA-----ATTGTCAGCCCAAGAGTGGTTGTAGAC----- 1623
 QY 87 GlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
 DB 1624 -----CCGGATAATGTATCTCTTAATGACAGAAACAGCA 1656
 QY 107 GlyTyrPheSerSerLysValSerLeuThrGluLysaspGlyAlaTyrThr----- 123
 DB 1657 GGACGC-----AGCAACTTTCAGAACGACGATGAATACACGGGATCCGGG 1701
 QY 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
 DB 1702 AATAGTGCAGACACCCCAACAGCAACAGAAAGAACACATTAACAACACCACTCTT 1761
 QY 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
 DB 1762 GAGAGTATATAAAAAAGGTACC-----TTTGTAAACATCACTGCTAATCAA 1809
 QY 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
 DB 1810 CGCATCTATGTCATAGTCCATTAATTTATCCAAATGGCAGCTTAACCTCTTTGGAGTGAG 1869
 QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
 DB 1870 GGTGCGAGCGGT-----GGCGCGCT----- 1890
 QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
 DB 1891 -----GAGATTAAACACGATATTACCACCGGTGATGATACACGAGGTGCAAAAC 1938
 QY 214 ValValValAspSerGlyArgProfile-----AlaPhe 224
 DB 1939 TTAACAATTTTACTACGCGCGGTGGTGTGATGTTTCATAAAAAATATCTCACTCGGGCGCAA 1998
 QY 225 GlyAspPheGluIleThrClyThrCln-----ArgTyrProGluGlnIle 239
 DB 1999 GGTAAACATAAACATTACAGCTTAACAGATATCCCTTTGAGAAAGAGCAACCAANGTC 2058
 QY 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuLeuLeuAsp 259
 DB 2059 ATTACAGGTCAAGGACATATTACCTCAGGCAAT----- 2091

QY 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerClyAlaSerValGlnAlaaspPhe 279
 DB 2092 -----CAAAAAAGGTTTAGATTTAATAATGCTCTCTCAACCGCAGCTGGC 2136
 QY 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
 DB 2137 ACGGAGTGCAG-----TTACCACTAAAGAACCAAT 2169
 QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
 DB 2170 AAATACGCTATCACAAATAAATTTGAAGGGGACTTTAAATAATTTTCAGGGAAGTGAACATC 2229
 QY 318 -----AlaTyrAsp-----TyrTyr 322
 DB 2230 TCAATGGTTTTACCTAAATAATCAAAAGTGGATATGATAAATCAAAAGGACCACTTACTGG 2289
 QY 323 AsnLeuPheAsnLys---GlyTyrIleClySerVal---ValTrpAspMetAspLysTyr 340
 DB 2290 AATTTAACCTCGAAAGTGGATATGATAAATCAAGGACCCCTCACTATTGACTCCAGA 2349
 QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 DB 2350 GGAAGCGATAGTGCAGGCACACATTACCAGCCTTATAATTTAAACGGT----- 2397
 QY 361 SerAsnValSerTyrAsnArgSerThrThrClnAsnLeuGluLysArgAlaPheSerGly 380
 DB 2398 -----ATATCATCTCAACAAAGACACTACCTTTAATCTTGACGAATGCA----- 2442
 QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
 DB 2443 -----AGATGCAACTTTGACATCAAGGCACCAATAGGATAAATAAGTAT 2487
 QY 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
 DB 2488 TCTAGTTTGAATTCACGATCATTTAATGGAACATTTTCAGTTTCGGGAGGGGGAGTGT 2547
 QY 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
 DB 2548 GATTTC-----ACACTTCGCTCATCC-----TCT 2574
 QY 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
 DB 2575 ACGTCCAAACCCCGGTGTAGTTATTAATTTCTAAATACTTTAATGTTTCAACAGGCTCA 2634
 QY 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
 DB 2635 AGTTTAAGATTTAAACCTTCAGGCTCAACAAAAAAGTGGCTTCTCAAT-AGAGAAAGATT 2693
 QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyClnAlaGlyTyr 488
 DB 2694 AACTTTAAATGCCAGCGGAGGACATACACTTTTGCAGATTGCAAGGACCGATGGAAT 2753
 QY 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
 DB 2754 GATTGTAAGGACATTTAGCCAAAAAACAATAACCTTTGAAGGAGGTAAAGATG---AG 2810
 QY 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
 DB 2811 GTTTGGCTCCAGGAAGCGGTAAACA-----GAATCGAAGGCATGTTACTATCAATAA 2864
 QY 524 oAsnGlySerValProGluArgAlaLeuLeuValGlySer-----LeuGluTyrGln 542
 DB 2865 CAACGCTAACGTC-----ACTCTTATCGGTTCGGATTTTGACAACCATCA 2909
 QY 542 nLeuProPheThr-----ArgThrLeuSerGln 551
 DB 2910 AAACCTTTAACTATTAAAAAAGATGTCATCATTAATAGCGGCACCTTACCCTCGGAGG 2969
 QY 551 yAlaValPheHisAspMetGly-----AspAlaAlaAsnPheLysArgPhe 567
 DB 2970 CAATATTGTCAATATATACCGGAAATCTTACCGTTGAAAGTAACGCTAATTTCAA----- 3024
 QY 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587

```
Db 3025 -----GCTATCAAAATTTTCAC 3041
Qy 587 rpeAspIleAlaTyrGlyrHisSerAspLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAATGTAGGCG---GGCTTTTTCAGACAAAGCAATTCAAATATTTCCATTGCCAA 3098
Qy 606 ---GlyThrArgPhe 609
Db 3099 AGGAGGGCTCGCTTT 3114

RESULT 26
US-08-469-880-5
; Sequence 5, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; FILING DATE: 16-MAR-1993
; APPLICATION DATA: US PCT/US93/02166
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:yg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-469-880-5 (1-9171)

Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
```

```
Db 1529 ATATCCCTAAACCCGGTGGT---GTGGAGAGCGTCGGGGCATGATTTATTTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluGluLeuAspLysGlu 86
Db 1588 AATGCA-----ATTTCTTCAACCCAAAGAGTGGTGTGTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGATATATGTATCTATTATTGCGAAGAACGCA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GGAGCG-----AGCAATACTTCAGACAGCATGAATACACGGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 AATAGTCCAGCACCCCAAAAGCAAAAGAACACACATTAACAACACACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATATAAAAAAGGTACC-----TTTGTAAACATCACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTATGCTCAATAGCTCCATTATTTATCCAAATGGCAGCTTAACCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTCCGAGCGGT-----GGCGGCGTT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GACATTAACAACGATATTACCACCGGTGATGATACCAGAGGTGCAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAACAATTTACTCAGCGCGGTGGTGTTCATATAAATATCTCACTCGGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAAACATAACATTACAGCTAAACAGATATCGCCTTGAGAAGGAGCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGGACTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGTTTTAGATTTAATAATGCTCTCTAAACGGCAGCTGGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 AGCGGACTGCAA-----TTCACCACCTAAAGAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AAATACGCTATCACAATAAATTTCAAGGGACTTTAAATATTTCAGGAAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCAATGGTTTTACTTAAATAAGTAGATATGATATAATTAAGAGGACGCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTTAACCTCGAAAGTGGATATGATAAATTCAAAGGACGCCCTCACTATTGACTCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTGCAGGCACACTTACCAGCGCTTATAATTTAAACCGT----- 2397
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
```

```

Db 2398 -----ATATCATTCACAAAGACACTACCTTTAATGTTGAACGAATGCA----- 2442
Qy 381 GlyIleTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCAACTTTTCACATCAAGGCAACCAATAGGATAAATAAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTTCAATTACGCATCATTTAATGGAACATTCAGTTTCGGGAGGGGAGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTTC-----ACACTTCTCGCCCTCATCC-----TCT 2574
Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AAGTCCAAACCCCGGTGTAGTTATTAATTTCTAATACTTTAATGTTTCAACAGGTCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 2635 AGTTTAAGATTTAATAACTTCAAGGCTCAACAAACCTGGCTTCTCAAT-AGAGAAAGATT 2693
Qy 469 PhePheThrProGluAsnLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTyr 488
Db 2694 AACTTTAATGCCCGGAGGACACATACACTTTTGCAGTTTGAAGGACCGCATGGAAT 2753
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGTAAGGCATTTGAGCAAAACAAACATAACCTTTGAAGGAGGTAAGATG--AG 2810
Qy 504 gSerGlyGlyAlaSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTGGCTCCAGGAAGCGGTACA-----GAATCGAAGGCAATGTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyrG 542
Db 2865 CAAGCGTAACGTC-----ACTCTTATCGGTTTCGGATTTTGACAACCATCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerG 551
Db 2910 AAAACCTTTTAACTATTAATAAAGATGTCATCATTAATAGCGGCAACCTTACCGCTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
Db 2970 CAATATTGTCAATATAGCGGGAATCTTACCGTTGAAGTAAGCTTAATTTCAAA----- 3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAATTTTCAAC 3041
Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTATGTAGC---GGCTTGTGTTGACAAACAAAGCAATTCAAATATTTCATTGCCAA 3098
Qy 606 ---GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114

RESULT 27
US-08-728-470-5
; Sequence 5, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF INVENTION: Of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia

```

```

; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Betkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-728-470-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-728-470-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
Db 1529 ATATCGCTAATAACCGCTGCTTTT-GTGAGACGTCGGGCGCATGATTTATTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLysGlu 86
Db 1588 AATGCA-----ATTGTTGAGCCCAAGAGTGTGTTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLys 106
Db 1624 -----CCGGATAATGCTATCTATTATTCAGAAACAGCA 1656
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
Db 1657 GGACGC-----AGCAATACTTCAGAGACGATGAATACACGGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 1702 ATAGTGCACGACCCCAACAAACGAAACAAAGAAACACACATTACAAACACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrp 162
Db 1762 GAGAGTATATAAAAAAAGGTACC-----TTTGTAACTACACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
Db 1810 CGCATCTGTCAATAGCTCCATTAATTATTCACATGGCAGCTTAACCTTTTGGAGTGAG 1869

```

QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 1870 GGTGGAGCGGT-----GGCGCGGT----- 1890
QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
Db 1891 -----GAGATTAAACAAGCATATTACCACCGGTGTATACCAGAGGTGCAAAAC 1938
QY 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
Db 1939 TTAAACAATTTACTCAGCGCGGTGGTGTATGTTCAATAAATATCTCACTCGGGGCGCAA 1998
QY 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
Db 1999 GGTAACATAAATACAGCTAAACAAGCATATCGCTTTTGAGAAAGAACCAACCAAGTC 2058
QY 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
Db 2059 ATTACAGGTCAAGGCACTATTACCTCAGGCAAT----- 2091
QY 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
Db 2092 -----CAAAAAGGTTTATAGATTAAATATCTCTCTAAACGCGCACTGGC 2136
QY 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
Db 2137 ACGGACTGCA-----TTCAACCACTAAAGAACCAAT 2169
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
Db 2170 AATACGCTATCACAAATAAATTTGAAGGACTTTAAATATTTCAAGGAAAGTGAACATC 2229
QY 318 -----AlaTyrAsp-----TyrTyr 322
Db 2230 TCATCGTTTACCTAAANATCAAGTCGATATGATAAATTTCAAGGACGCACTACTGG 2289
QY 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
Db 2290 AATTTAACTCGAAAGTGATGATATAAATTTCAAGGACCGCTCACTATTGACTCCAGA 2349
QY 341 GluThrThrLeuAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
Db 2350 GGAAGCGATAGTCAGGCACACTTACCAGCTTATTAATTTAAACGGT----- 2397
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 2398 -----ATATCATTCACAAAGACACTACCTTTAATGTTCAACGAATGCA----- 2442
QY 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
Db 2443 -----AGAGTCACTTTGACATCAAGGCACCAATAGGATTAATAGTAT 2487
QY 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 2488 TCTAGTTGAATTACGCATCATTTAATGGAAACATTTTCAGTTTCGGAGGGGGAGTGT 2547
QY 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTCT-----ACACTTCTCGCTCACTC-----TCT 2574
QY 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
Db 2575 AAGGTCACAAACCCGGGTGTAGTTATATATTTCTAAATATTTAATGTTTCAACAGGGTCA 2634
QY 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
Db 2635 AGTTTAAGATTTTAAACTTTCAGGCTCAACAAAAAAGTGGCTTCTCAAT-AGAGAAAGATTT 2693
QY 469 PhePheThrProGluAsnLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTyr 488
Db 2694 AACTTTAAATGCCCGGAGGCAACATAACACTTTTTCGAAGTTGAAGGACCGCATGGAT 2753

QY 488 r-----ThrValAlaArgAspAsnAlaAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATTGGTAAAGCATTTAGCCAAAAAACAATTAACCTTTGAAGGAGGTAAAGATG---AG 2810
QY 504 gSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGGAAAGCGGTAAACA-----GAAATCGAAGGCAATGTTACTATCAATAA 2864
QY 524 OAsnGlySerValLeuProGluArgAlaLeuLeuValGlyLysSer-----LeuCluTyrG 542
Db 2865 CAACGCTAACGTC-----ACTCTTATCGGTTCCGATTTTGACAAACCATCA 2909
QY 542 nLeuProPheThr-----ArgThrLeuSerG 551
Db 2910 AAAACCTTTTAACATTAATAAAAGATGTTCATTAATAAGCGCAACCTTACCCTGGAGG 2969
QY 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPhelysArgMe 567
Db 2970 CAATATTGCTAATATAGCGGGAATCTTACCGTTGAAGGTAAACGCTAATTTCAAA----- 3024
QY 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAAATTTTCAC 3041
QY 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTATGTAGGC---GGCTTGTGTGACACAAAGGCAATTCAAATATTTCCATGCCAA 3098
QY 606 ---GlyThrArgPhe 609
Db 3099 AGGAGGGGCTCGCTTT 3114
RESULT 28
US-08-617-697-5
; Sequence 5, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF INVENTIONS: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

: LENGTH: 9171 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 US-08-617-697-5

Alignment Scores:

Pred. No.: 0.23 Length: 9171
 Score: 114.00 Matches: 138
 Percent Similarity: 32.82% Conservatives: 74
 Best Local Similarity: 21.36% Mismatches: 235
 Query Match: 3.59% Indels: 201
 DB: 2 Gaps: 33

US-09-857-669-2 (1-609) x US-08-617-697-5 (1-9171)

Qy 47 valLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
 Db 1529 ATATCGCTAAACCGGTGTTT-GTGGAGACGTCGGGGCATGATTATTCATCAAGAC 1587
 Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGlu 86
 Db 1588 AATGCA-----ATTGTTGACGCCAAAGAGTGGTTGTAGAC----- 1623
 Qy 87 GlnThrGlyPheLeuAlaGluLalaProAspAsnValLysThrMetLeuArgSerLys 106
 Db 1624 -----CCGGATAATGTTATTAATCGCAGAAACGCA 1656
 Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
 Db 1657 GGACGC-----AGCAATACTTCAGAACGCGATGAATACGGGATCCGG 1701
 Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
 Db 1702 ANTAGTGCAGCAGCCCAAGAACAAAGAAAGAACAAACATTAACAAACACACTCTT 1761
 Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
 Db 1762 GACAGTATACTAAAGGTTACC-----TTTGTTAAACATCACTGCTAATCAA 1809
 Qy 163 GlnGlnProValClySerAspPheAspGlnAspSer-----TrpGluAsn 177
 Db 1810 CGCATCTATGTCAATAGCTCCATTAATTTATCCATGGCAGCTTAACCTCTTGGAGTAG 1869
 Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
 Db 1870 GGTCCGCGCGT-----GGCGCGCT----- 1890
 Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
 Db 1891 -----GAGATTAAACACGATATTACCACCGGTGATGATACCGAGGTGCAGAAC 1938
 Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
 Db 1939 TTAACAATTTACTACGCGGTGGTGTGATGTTCAATAAAATATCTCACTCGGGCGCA 1998
 Qy 225 GlyAspPheGluIleThrClyThrGln-----ArgTyrProGluGlnIle 239
 Db 1999 GGTAACATAAATTAACAGCTAAACAGATATCGCCTTTGAGAAAGGAGAACCAAGTC 2058
 Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
 Db 2059 ATTACAGGTCAAGGGACTATTACCTCAGGCAAT----- 2091
 Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
 Db 2092 -----CAAAAGAGTTTTAGATTTAATATGTTCTCTAAACGGCATCGGC 2136
 Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
 Db 2137 AGCGGACTGCAA-----TTCCACCACTAAAGAACCAAT 2169

Qy 300 LysLeuGluThrGlyLeArgLeuAspSerGluTyrClyLeuGlyGlyLysIle----- 317
 Db 2170 AAATACGCTATCACAAATAAATTTGAAGGGACTTTAAATATTTTCAGGGAAGTGAACATC 2229
 Qy 318 -----AlaTyrAsp-----TyrTyr 322
 Db 2230 TCAATGGTTTTACCTAAATAATGAAGTGGATATGATAAATTCAAAGACGCACTTACTGG 2289
 Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
 Db 2290 AATTTAACCTCGAAAGTGGATATGATAAATTCAAAGACCGCCTCACTATTGACTCCAGA 2349
 Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
 Db 2350 GGAAGCGATAGTGCAGCACACTTACCCAGCCTTATATTTAAACGGT----- 2397
 Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
 Db 2398 -----ATATCATTCACAAAGACACTACCTTTAATCTTGACGAATGCA----- 2442
 Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
 Db 2443 -----AGAGTCAACTTTGACATCAAGGCACCAATAGGATAAATAAGTAT 2487
 Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
 Db 2488 TCTAGTTTCAATTACCCATCATTTAATGGAACATTTCACTTTCCGGAGGGGGAGTGT 2547
 Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
 Db 2548 GATTTC-----ACACTTCGCCCTCATCC-----TCT 2574
 Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTyrLeuAspGlyLysIleGlyThr 448
 Db 2575 AGCTCCAAACCCCGGTGTAGTTATAAATTTCTAAATCTTAACTTTATGTTTCACAGGGTCA 2634
 Qy 449 ThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr 468
 Db 2635 AGTTTAAGATTTAAACTTCAGGCTCAACAAACAACTGCTTCTCAAT-AGAGAAAGATTT 2693
 Qy 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTyr 488
 Db 2694 AACTTTAAATGCCAGCGGACCACTAATCACTTTTCAAGTTGAAGGACCACCATGGAAT 2753
 Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
 Db 2754 GATTGTAAGGCGATTGAGCCAAACAAACATAACCTTTCAAGGAGGTAAAGATG---AG 2810
 Qy 504 qSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPr 524
 Db 2811 GTTTGGCTCCAGAAAGCGGTAAAC-----GAAATCGAAGGCANTGTTACTATCAATRA 2864
 Qy 524 oAsnGlySerValLeuProGluArgAlaLeuValGlySer-----LeuGluTyrG 542
 Db 2865 CAACGCTAACGTC-----ACTCTATCGTTCGGATTTTGACAAACCATCA 2909
 Qy 542 nLeuProPheThr-----ArgThrLeuSerG 551
 Db 2910 AAACCTTTAACTATTAAAAAGATGTCATCATTAATAGCGCAACCTTACCCTCGAGG 2969
 Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPheLysArgMe 567
 Db 2970 CAATATTGTCAATATAGCCGGAATCTTACCGTTGAAAGTAAACGCTAATTTTCAAA----- 3024
 Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPhe 587
 Db 3025 -----CCTATCACAATTTTCAC 3041
 Qy 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
 Db 3042 TTTTAATGTAGGC---GGCTTGTGACAAACAAAGGCAATTCAAATATTTCCATTTGCCAA 3098
 Qy 606 ---GlyThrArgPhe 609

```

Db 3099 AGGAGGGCTCGCTTT 3114
|||||
RESULT 29
US-08-719-641-5
; Sequence 5, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-719-641-5

Alignment Scores:
Pred. No.: 0.23 Length: 9171
Score: 114.00 Matches: 138
Percent Similarity: 32.82% Conservative: 74
Best Local Similarity: 21.36% Mismatches: 235
Query Match: 3.59% Indels: 201
DB: 4 Gaps: 33

US-09-857-669-2 (1-609) x US-08-719-641-5 (1-9171)
Qy 47 ValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAsp 66
::: |||||||
Db 1529 ATATCGCTAAACCGGTGCTTTT-GTGGAGACGTCGGGGCATGATTATTCATCAAGAC 1587
Qy 67 MetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGlu 86
::: |||||||
Db 1588 AATGCA-----ATTGTCACGCCCAAGAGTGGTGTAGAC----- 1623
Qy 87 GlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLysSerLys 106
::: |||||||

```

```

Db 1624 -----CCGATATATGATCTATTATTAATCGAAGACGCA 1656
|||||
Qy 107 GlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr----- 123
||| ||| ||| ||| |||
Db 1657 GGACGC-----AGCAATACTTTCAGAAGACGATCAATACACGGGATCCGGG 1701
Qy 124 ---ValHisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaLeu 142
||||| ||| ||| ||| |||
Db 1702 AATAGTCCGACGACCCCAACAAAGAAAGACACATTTAACAAACACACTCTT 1761
Qy 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
||||| ||| ||| ||| |||
Db 1762 GAGAGTATATACTAAAAAAGGTACC-----TTTGTTAACATCACTGCTAATCAA 1809
Qy 163 GlnGlnProValGlySerAspPheAspGlnAspSer-----TrpGluAsn 177
::: ||| ||| ||| ||| |||
Db 1810 CGCATCTATGTCATAGTCCATTAATTTATCCAAATGGCAGCTTAACCTTTGGAGTGAG 1869
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
::: ||| ||| ||| ||| |||
Db 1870 GCTCGGAGGGGT-----GGCGGCGTT----- 1890
Qy 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAsp-----LeuAsn 213
::: ||| ||| ||| ||| |||
Db 1891 -----GAGATTAAACACGATATTACCACCGGTGATGATACCAGAGGTGCAAC 1938
Qy 214 ValValValAspSerGlyArgProIle-----AlaPhe 224
::: ||| ||| ||| ||| |||
Db 1939 TTAACAAATTTACTCAGCGGCGTGGTGTGATGTTCAATAAATATCTCACTCGGGGCGCAA 1998
Qy 225 GlyAspPheGluIleThrGlyThrGln-----ArgTyrProGluGlnIle 239
||| ||| ||| ||| ||| |||
Db 1999 GGTACATTAACATTAACACCTAAACACATATCGCCTTTGAGAAAGACCAACCAAGTC 2058
Qy 240 ValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAsp 259
||| ||| ||| ||| ||| |||
Db 2059 ATTACAGGTCAAGGGCACTATTACCTCAGGCAAT----- 2091
Qy 260 PheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPhe 279
||| ||| ||| ||| ||| |||
Db 2092 -----CAAAAAGGTTTAGATTAAATGCTCTCTTAACGGCCTGCGC 2136
Qy 280 AspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHis 299
||||| ||| ||| ||| ||| |||
Db 2137 AGCGGACTGCAA-----TTCACCACTAAAGAACAACCAAT 2169
Qy 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIle----- 317
||| ||| ||| ||| ||| |||
Db 2170 AATACGCTATCACAAATAAATTTCAGAGGACTTTAAATATTTCAGGGAAGTGAACATC 2229
Qy 318 -----AlaTyrAsp-----TyrTyr 322
||| ||| ||| ||| ||| |||
Db 2230 TCAATGGTTTACCTAAAATGAAAGTGGATATGATAAATTCAAAGCAGCCACTTACTGG 2289
Qy 323 AsnLeuPheAsnLys---GlyTyrIleGlySerVal---ValTrpAspMetAspLysTyr 340
||| ||| ||| ||| ||| |||
Db 2290 AATTTAACCTCGAAAGTGGATATGATAAATTCAAAGACGCCCTCACTATTGACTCCAGA 2349
Qy 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
::: ||| ||| ||| ||| ||| |||
Db 2350 GGAAGCGATAGTGCAGGCACACTTACCAGCGCTTATATTTAAACGGT----- 2397
Qy 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
||| ||| ||| ||| ||| ||| |||
Db 2398 -----ATATCATCAACACACACACTACCTTTAATGTTGAACGAATGCA----- 2442
Qy 381 GlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly----- 396
||| ||| ||| ||| ||| ||| |||
Db 2443 -----AGACTCACTTTGACATCAAGGCACCAATAGGAGTAATAATAGTAT 2487
Qy 397 -----AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
||| ||| ||| ||| ||| |||

```

```
Db 2488 TCTAGTTTCAATTACGCATCATTTAATGGAACATTTTCAGTTTCGGGAGGGGGAGTGT 2547
Qy 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 2548 GATTTTC-----ACACTTCTCGCTCATCC-----TCT 2574
Qy 432 AsnValLeuHisProGlu-----AsnGlyHisTrpLeuAspGlyLysIleGlyThr 448
Db 2575 ACAGTCCAAACCCCGGTGCTAGTTATTAATCTTAATATCTTAATGTTTCAACAGGGTCA 2634
Qy 449 ThrLeuGlyThrPheLeuSerSerHisAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 468
Db 2635 ACTTTAAGATTATTAACCTTCAGGCTCAACAAAACCTGCTCTCAAT-AGAGAAAGATT 2693
Qy 469 PhePheThrProGluAsnLysLeuGlyThrPheIle-IleArgGlyGlnAlaGlyTy 488
Db 2694 AACTTTAATGCGCCGAGGACACATTAACACITTTGCAAGTTGAAGGACCGATGGAAT 2753
Qy 488 r-----ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheAr 504
Db 2754 GATGCGTAAGGCATTTAGCCAAAACATAACCTTTGAAGGAGGTAAGATG---AG 2810
Qy 504 gSerGlyGlyAlaSerSerValArgGlyTyTrpGluLeuAspSerIleGlyLeuAlaGlyPr 524
Db 2811 GTTTGGCTCCAGGAAGCGGTAAAC-----GAAATCGAAGCAATGTTACTATCAATAA 2864
Qy 524 oAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer-----LeuGluTyG 542
Db 2865 CAACGCGTACGTC-----ACTCTTATCGGTTCGGATTTTGACAACCATCA 2909
Qy 542 nLeuProPheThr-----ArgThrLeuSerG 551
Db 2910 AANAACCTTTAATATAAAGAGATGTCATCATTAATAGCGGCAACCTTACCGCTGGAGG 2969
Qy 551 yAlaValPheHisAspMetGly-----AspAlaAlaAlaAsnPhelysArgMe 567
Db 2970 CAATATTGTCAATATAGCGGAAATCTTACCGTTGAAAGTAAGCGTAANTTTCAA----- 3024
Qy 567 tLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSe 587
Db 3025 -----GCTATCACAAANTTTCAC 3041
Qy 587 rPheAspIleAlaTyrglyHisSerAspLysLysIleArgTrpHisIleSerLeu----- 605
Db 3042 TTTTAATGTAGGC---GCGTGTGTTGACAAACAAAGCAATTTCAAAATTTTCCATTGCCAA 3098
Qy 606 ---GlyThrArgPhe 609
Db 3099 AGCAGGGGCTCGCTT 3114
```

RESULT 30

US-08-021-601-9
Sequence 9, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
METHODS OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSEE: Needle & Rosenberg, P. C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
US-08-021-601-9

Alignment Scores:

Pred. No.: 0.0189 Length: 1524
Score: 112.50 Matches: 112
Percent Similarity: 34.80% Conservative: 78
Best Local Similarity: 20.51% Mismatches: 189
Query Match: 3.55% Indels: 167
DB: 1 Gaps: 28

US-09-857-669-2 (1-609) x US-08-021-601-9 (1-1524)

```
Qy 38 LysAsnLysSerProAspThrGluSerValLysLysLysProLysPheProValArgIle 57
Db 40 AAAGAGAAAATAAAGATGAGAAT-----AAGAGAAAAGATCAAGAACCAAT 87
Qy 58 AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGln 77
Db 88 AAACACAGAGAGAGCATTAAAGGAATCATG---AAACACATTTGTAANAATGAAGTA 144
Qy 78 GlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAsp 97
Db 145 AAAGGGGAGGAGCTGTGTAAAAAAGAGGACAGACAGAAAAGCTACTTGAGAAAGTACCATCT 204
Qy 98 AsnValLysThrMetLeuArgSerLysGlyTyTrpPheSerSerLysValSerLeuThrGlu 117
Db 205 GATGTTTATAGATGTATTAAGCAATTGGA-----CGAAAGATATATATTGTG--- 252
Qy 118 LysAspGlyAlaTyThrValHisIleThr-----ProGlyProArgThrLys 133
Db 253 ---GATGTTGATATTACAAAACATATATCTTTAGAACCATTTATCTGAAGATAAGAAAAA 309
Qy 134 IleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerSerGlyAsnLeuAlaGlu 153
Db 310 ATAAAGAC-----ATTTATGGG-----AAAGATCTTTATTATACATGA 348
Qy 154 TyrTrpArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerPheAspGlnAsp 173
Db 349 CATTATGTATATGCAAAAAGAGGATATGAACCCGTACTTGTAAATCCCAATCTTCGGAAGAT 408
Qy 174 SerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyTrpProLeu 193
Db 409 TATGTAGAAAATCTGAAAGGCA-----432
Qy 194 AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsn 213
Db 433 -----CTGAAC 438
Qy 214 ValValAlaAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGln 233
Db -----CTGAAC 438
```

Db 439 GTTTATTATCAATAGGTAAAGATATTATCA----- 468

Qy 234 ArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAsp 253
||||:||||| :||| :|||||

Db 469 -----AGGATATTTAAAGTAAAT-----AATCAACCATCAG 504

Qy 254 -----LeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSer 271
||||:||||| :||| :||| :||| :||| :|||

Db 505 AATTTTATAGATGATTAATACCATTAATAATGATGATTCAGATGCACAA----- 558

Qy 272 GlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysVal 291
||| :||| :||| :||| :||| :|||

Db 559 -----GATCTTTA----- 567

Qy 292 SerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyr 311
:||||:||||| :||| :||| :||| :||| :|||

Db 568 TTTACTAATCAGTAAAGCAACATCCACAGACTTTCTGTGAGATTTCTTGGACAAAT 627

Qy 312 GlyLeuGlyLysIleAlaTyrTyrTyrAsnLeuPheAsnLysGlyTyrIleGly 331
:||||:||||| :||| :||| :||| :||| :|||

Db 628 -----AGCAATGAGTACAAAGATATTTCGAAAGCTTTT----- 663

Qy 332 SerValIleTyrAspMetAspLysTyrGluThrLeuAlaGlyIleSerGlnPro 351
:||||:||||| :||| :||| :||| :||| :|||

Db 664 ---GCATATTATATCAGCCACAGCATCGTGTCTTTACAGCTTTTATGCACCGAAGCT 720

Qy 352 ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln 371
||||| :||| :||| :||| :||| :|||

Db 721 TTTAATTACATGATTAATTAACCAACAAAGAAATAAATCTAACGCTGCGGCACAGCC 780

Qy 372 AsnLeuGluLys-----ArgAlaPheSer 379
:||||:||||| :||| :||| :||| :||| :|||

Db 781 GAGTGTGAGCTGACCTGCGCGTCCGCGGTGAATGCGCGCGCGGACAGC 840

Qy 380 GlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeu----- 395

Db 841 GGC-----GACCGCTCTGGAGCGCAACTAT 867

Qy 396 -----GlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
:||||:||||| :||| :||| :||| :||| :|||

Db 868 CCCACTGGCGCGAGTCTCGCGCGCGCGCGCGCGCTC----- 906

Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrp-----LysArgGlnLeuLeuAsn 431
||| :||||:||||| :||| :||| :||| :|||

Db 907 -----AGTTACAGCACCGCGCGACGAGACTGGAGGTGGAGCGGCTCCAGCGC 960

Qy 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
:| :||| :||| :||| :||| :||| :|||

Db 961 CACCGCAACTGGAGAGCGCGCTATGTTCTGCGCTACCGCGCACCTTCCTCGAA 1020

Qy 452 ThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPheThr 471
:||||:||||| :||| :||| :||| :||| :|||

Db 1021 -----CGCGCGCAAGCATGCTTCGCGCGGTGCGCGC----- 1056

Qy 472 ProGluAsnLysLysLeuGlyThr-----PheIleIleArgGlyGln----- 485
:||||:||||| :||| :||| :||| :||| :|||

Db 1057 ---CGCAGCGCACTCGACGCGATCTGCGCGGTTCATATACGCGCGCATCGCGC 1113

Qy 486 AlaGlyThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer 505
||| :||||:||||| :||| :||| :||| :|||

Db 1114 CTGCGCTACGGCTACCGCGAGGCCAGCAACCGACGCGCGC---CGGATCCGCAAC 1170

Qy 506 GlyGly-----AlaSerSerValArgGlyTyrGluLeuAspSer 518
||| :||||:||||| :||| :||| :||| :|||

Db 1171 GGTGCGCTGCTCGGTCTATGTGCGCGCTCGAGCTGCGCGCTCTACCGCACGCG 1230

Qy 519 IleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySer 538
:| :||||:||||| :||| :||| :||| :|||

Db 1231 CTGACCTGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1281

Qy 539 LeuGlyTyrGlnLeuPro 544
:| :||||:||||| :||| :||| :||| :|||

Db 1282 -----CATCCGCTGCCG 1293

RESULT 31

US-08-082-849B-9
: Sequence 9, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Atora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: TITLE OF INVENTION: Related Methods
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082.849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021.601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31.677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1524 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus anthracis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1524
: OTHER INFORMATION: /product= "LF(1-254)--TR--PE(362-613)"
: US-08-082-849B-9

Alignment Scores:
Pred. No.: 0.0189 Length: 1524
Score: 112.50 Matches: 112
Percent Similarity: 34.80% Conservative: 78
Best Local Similarity: 20.51% Mismatches: 189
Query Match: 3.55% Indels: 167
DB: 1 Gaps: 28

US-09-857-669-2 (1-609) x US-08-082-849B-9 (1-1524)

Qy 38 LysAsnLysSerProAspThrClnSerValLysLeuLysProLysPheProValArgIle 57
:||||:||||| :||| :||| :||| :||| :|||

Db 40 AAGAGAAAATAAAGATGAGAT-----AAGAGAAAAGATGAGAACGAAT 87

Qy 58 AspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGln 77
:||||:||||| :||| :||| :||| :||| :|||

Db 88 AAAACACAGGAAGAGCATTTAAGGAATCATG---AAACACATTTGTAATAATAGAAGTA 144

STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10014

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,124A
FILING DATE: 23-DEC-1994
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,254
FILING DATE: 20-SEP-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/973,336
FILING DATE: 05-NOV-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,187
FILING DATE: 23-AUG-1990

ATTORNEY/AGENT INFORMATION:
NAME: Gallagher, Thomas C.
REGISTRATION NUMBER: 37,066
REFERENCE/DOCKET NUMBER: SPA-1-PDC

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 645-1405
TELEFAX: (212) 645-2054

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3286 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:
ORGANISM: Neisseria gonorrhoeae
STRAIN: FA19
FEATURE:
NAME/KEY: CDS
LOCATION: 406..3150

US-08-363-124A-1

```
QY 389 aGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySe 409
Db 1611 GCACACCAAAAGCCCTACGTTTGGAAATATGTC-----TATACCAATGCC 1655
QY 409 rAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLe 429
Db 1656 CGATAAAGACACTTGGCGGGATATATGCCGCCCTC-----TCITACGACCGGCAG-- 1704
QY 429 uLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrTh 449
Db 1705 -----GGCATCGGTTTGGCAACACCATTTTCACGACGAC 1736
QY 449 rLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPh 469
Db 1737 GCACGTCTTCGGCGGAGGTTCCGCAAAATATTCGCGTCCGAGTCCGCAACGCGGTTTC 1796
QY 469 ePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly-----484
Db 1797 CTATTACAAATCCGACCGC-----GTGATTTACGGGGAACCAATTAAGCT 1841
QY 485 ----GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPh 503
Db 1842 CTTGACGGCGGCAATTCACAAAATCTTCGATACGCGCAAAATCCGCCACACCTGAGCCT 1901
QY 503 eaArgSerGly-----GlyAla-----Se 509
Db 1902 GAATCTCGTTACGACCGCTTCGGGCTCTAATCTCCGCCATCAGGATTATTATATCAAG 1961
QY 509 rSerValArgGlyTyrGluLeuAsp-----517
Db 1962 TCGCAACCGCGCTATTCTGTGAAACCGCCCTCAAAACCGCAACGCAAAACCGCC 2021
QY 518 -----SerIleGlyLeuAlaGlyProAsnGlySe 527
Db 2022 CAACGGCAGAGAAAAGATCCCTATTGGTCAGCATAGGC-----AGGGGAAA 2069
QY 527 rValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrAr 547
Db 2070 TGTGCTAGGAGGCAAAATCTGCTCTTTGGCAACAATCTATATACGACTGCACGCGCG 2129
QY 547 gThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPhelysArgMe 567
Db 2130 CAGCATCAACGGCAAAAGCTATTAC-----GC 2156
QY 567 tLysLeuHisGlySerGlyLeuGlyValArgTyrPheSerProLeuAlaProPheSe 587
Db 2157 GCGGTTCGGGCAAAATGTCCTTTGGGC---AGGTGGCGGATGTCGGCGCGGCTTGGC 2213
QY 587 rPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGlyTh 607
Db 2214 CTACGACTACCCGACGACCATTCGGAGAC-----GGCAGCGTTTCACCGGCAC 2264
QY 607 r 607
Db 2265 G 2265
RESULT 34
US-08-565-386-10
; Sequence 10, Application US/08565386
; Patent No. 5741697
; GENERAL INFORMATION:
; APPLICANT: Bavoil, Patrik M.
; APPLICANT: Hsia, Ru-ching
; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP,
; CITY: Roches
; STREET: Clinton Square, P.O. Box 1051
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/565,386
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timlan, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 376/60040
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1636
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-565-386-10
Alignment Scores:
Pred. No.: 0.0314 Length: 1791
Score: 111.50 Matches: 107
Percent Similarity: 35.60% Conservatve: 55
Best Local Similarity: 23.52% Mismatches: 214
Query Match: 3.51% Indels: 79
DB: 1 Gaps: 22
US-09-857-669-2 (1-609) x US-08-565-386-10 (1-1791)
QY 14 PhePhePheProHisAlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGly 33
Db 352 TATTTTGTCTCAAGTAACCTTCTCCTGATGGAGGTTACCGCAGAGAAATCTATTATGAC 411
QY 34 PheAlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhe 53
Db 412 TATTTTGGTCTACCTACTAAAGTTGCAAAATATTCGGCAT---CAAGTTTTCGCAATTAAGA 468
QY 54 ProValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPro 73
Db 469 GCATATATTTGATTTAATGATATATATAGTTAATGCTGCGCAATAGCGATATGATTAT 528
QY 74 LeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGlu 93
Db 529 GTTTGGACAGGAGATGCTGATCTCAAGGTTGAT---CCGACTACTGGAGAAGAAATCTCAA 585
QY 94 Glu-----AlaProAspAsnValLysThrMetLeuArgSerLys-----GlyTyr 108
Db 586 GAGGATGATCGACGTTCTCTATATATAGTTAATGCTGCGCAATAGCGATATGATTAT 645
QY 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 646 TTTACTTCTGCTCTCTCTGTTTGCAGAAAGTCTCTCTGTTGGAATAGTATTACAGGT 705
QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIle-----LeuGlyAspIle 145
Db 706 GGAGATTACGAGGACGCTCCAGTTTCATGTTTACGATTAGATCTATTTTGGATGATTCT 765
QY 146 LeuSerAspGly---AsnLeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnGln 164
Db 766 TCTGATGATCAGCTTTAGTTTGGTGTCTTCTTATGTAACGCTTCACAGAAATGG----- 819
QY 165 ProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184
Db 820 -----TTTACTGCGCAT---GTCCTTTGACTTCTGGAATGGGT 855
QY 185 AlaVal-----ThrArgLysGlyTyrProLeu-----AlaLys 195
Db 856 AGTGTTCTCTGTTGGTACAACTGGTAATTTTCTTATGATTAATCTTGTGTATCCATCTTAT 915
QY 196 LeuGlyAsnThrArgAlaValAlaAsnProAspThrAlaThrValAspLeuAsnValVal 215
Db 111 111 111
```

Db 916 TTTGGTAGACACTGTTGGCCAAACCTGGTAGTCCATCTTCTTCT----- 957
QY 216 ValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyr 235
Db 958 -----TCTACTCCGCCCTTTGGTTAAAGGGTGATTTCTCTGTT-----TAT 996
QY 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 997 GTTGAT-----TTAGCGGCTTCATCTTCAGTTACGATTAATTCGCTCGT 1041
QY 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGly----- 272
Db 1042 AATGGCATTACTTCCAAACAGCTGTTGAGAAGAGTCTGCTTTATGAGATAGATATGTT 1101
QY 273 AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp-----ArgValProVal 289
Db 1102 GAATCTGTTCAAGGTCATTTGGCGCTCATCTTGGTGATTCGCTCAGCGACCAATC 1161
QY 290 -----LysValSerValThrGluValLysArgHisLysLeuGlu 302
Db 1162 TATTTAGTGGATCAAGTCTTATGTTCTGTTAACTCTGATACAGAAATTC 1215
QY 303 ThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr----- 319
Db 1216 -----TCTACAGATTTCAGTTCTCCTCAAGGAAATCTTCTGCTTATGCATTATCT 1266
QY 320 ---AspTyrTyrAsnLeuPheAsnLysGlyTyrIle-----GlySerValValTrpAsp 336
Db 1267 ACAGATACATAACATTTCTTACGAAGTCTTTGTTGAGCATGGTTTCTTATAGCTCTT 1326
QY 337 MetAsp-----LysTyrGluThrThrLeuAlaAlaGlyIleSerGlnPro 351
Db 1327 CTTTCAGCTACAGCGGATTTAACTTATCAGCAAGGTTTAGAGCGTCAGTGGTCAAGATT 1386
QY 352 ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrGln 371
Db 1387 AGTCGGTAT--GATTATATTAGCCCTACTTTGCTCATTTGGGAGAGCGGCTGTTAT 1443
QY 372 AsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIle 391
Db 1444 AATAAGAGATTTATTGGCAATCAGATACTGTTATGATCCTAGTGGTTCGGGTTAAT 1503
QY 392 AspAlaArgLeuGly--AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp 410
Db 1504 GATGTCCTTTTGGTTATCAAGAGCGTTATGCTGAGTATCGTTATAGCGCTTCGAAGGTT 1563
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrp 425
Db 1564 ACTGGATTATTTAGATCTCAAGCTACAGGTACTCTAGATTCTTGG 1608

RESULT 35
US-09-090-793-13
; Sequence 13, Application US/09090793
; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Calgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; of polypeptide-like synthesis genes in plants
; FILE REFERENCE: CGNE 131.01US
; CURRENT APPLICATION NUMBER: US/09/090,793
; EARLIER FILING DATE: 1998-06-04
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-13
Alignment Scores: 1.41 Length: 19227
Pred. No.:

Score: 111 50 Matches: 77
Percent Similarity: 35.29% Conservative: 55
Best Local Similarity: 20.59% Mismatches: 137
Query Match: 3.51% Indels: 105
DB: 3 Gaps: 15
US-09-857-669-2 (1-609) x US-09-090-793-13 (1-19227)
QY 76 ThrGlnGlnGlnGluValIleAspLysGlnGlnThrGlyPheLeuAlaGluGluAla 95
Db 9162 ACCGCATTCACCAAGCAGTAGCATATACCTCAGACGCGTTCGTTGTTAGAGTGTGACGA 9221
QY 96 ---ProAspAsnValLysThrMetLeuArgSerLysGly-----TyrPheSer 110
Db 9222 GTCGCTGATTCGCAATCGCATTTGCTGAAAGCCAAGGTTTAATGCTGCTTATCATCAT 9281
QY 111 SerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrProGlyPro 130
Db 9282 ACCAAACTTTCATCTGCTATTAAGCAGTCCCGTGTGACTGGTGANGCGCGGTGT 9341
QY 131 ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn 150
Db 9342 TTTTCACAGCTCGCAGGTTTATGAAATGTCTAATTCGT----- 9380
QY 151 LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySerAspPhe 170
Db 9381 TTACATCAACGTTATATTCGCGGATTAAGATTGGCAACACCG-----AGTGACAAT 9434
QY 171 AspGlnAspSerTrpGluAsnSerLysThrSerValIleGlyAlaValThrArgLysGly 190
Db 9435 CAAATGTCACGCTGGCGCAATTC-----CCATTCTATATGCTGTAGATGCTCAGCTTGG 9491
QY 191 TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
Db 9492 TTCCACATCTGATGCTGCTGCACACATTCGCGCTTATGTTGTGCTGCTGCACACG 9551
QY 211 AspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
Db 9552 TATTGTTCATATCTTTTACAA----- 9572
QY 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThr 250
Db 9573 -----GAAACGCTCTTACAGAACTTGT---TTGAAGAAACAGTCT 9611
QY 251 ProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGln----- 266
Db 9612 TTGCAAGATAATGACTTAACTCAAGCAAGCTTCAGACTCTTGAACAAACAAATCCAGTA 9671
QY 267 -----AsnGlyHisTyrSerGly----- 272
Db 9672 CCTGATCTCGGCACATAATGTTACTTTCCATCGACGCGAGTTAGCATTAATCATAGTACAA 9731
QY 273 -----AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValPro 288
Db 9732 GGTAATGACCAAGCACACAAATACGCTGTGTAATTAGAAACTATTACAGGCGAGTTAAGTACT 9791
QY 289 ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp 308
Db 9792 ACTGGCATAGTACTACTATCAGTATTAAA----- 9818
QY 309 SerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyr-----AsnLeuPheAsn 326
Db 9819 -----CAGATCGCAGCAGACTGTTATGCCGTAATGATGATGATAAC 9857
QY 327 LysGlyTyr-----IleGlySerValValTrpAspMetAspLysTyrGluThr 342
Db 9858 AAAGCCTATAGCCGAGTCTTATTCGCCGAGACTGCTCAGAGACTGCTAGAGAAATAACC 9917
QY 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr----- 354
Db 9918 TTGGCGTTTGTGCTATGCTAGCTGTTTAAAGATGCTAAAGAAATGAAACCCCG 9977
QY 355 ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg-----SerThrThrGlnAsn 372

```

Db 9978 AAGGCGAGTATTTTACCGCGCAGCCTGCAAAATAACAGAGCTCTTAACAGCACAGAAAT 10037
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAsp 392
Db 10038 GGTGTACACCTTCATGATACCCAGGT----- 10061
Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIle 406
Db 10062 -----ATTGGTGCTACATATGTTGGTTAGGCGGTGATCTA 10097

RESULT 36
US-09-090-793-12
; Sequence 12, Application US/09090793
; Patent No. 6140486
; GENERAL INFORMATION:
; APPLICANT: Celgene, LLC
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants
; FILE REFERENCE: CGNE.131.01US
; CURRENT APPLICATION NUMBER: US/09/090,793
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 40138
; TYPE: DNA
; ORGANISM: Vibrio marinus
US-09-090-793-12

Alignment Scores:
Pred. No.: 4.58 Length: 40138
Score: 111.50 Matches: 77
Percent Similarity: 35.29% Conservative: 55
Best Local Similarity: 20.59% Mismatches: 137
Query Match: 3.51% Indels: 105
DB: 3 Gaps: 15

US-09-857-669-2 (1-609) x US-09-090-793-12 (1-40138)
Qy 76 ThrGlnGlnGluGluValLeuAspLysGlyGlnThrGlyPheLeuAlaGluGluAla 95
Db 26145 ACCGATTGCACCAAGCTAGCATACGACGAGCGAGTGGTTGTTAGAAAGTGTACGCA 26204
Qy 96 ---ProAspAsnValLysThrMetLeuArgSerLysGly-----TyrPheSer 110
Db 26205 GTCGCTGATTGCGCAATCGCATTTGCTGAAAGCCAAAGGTTTAATGCTGCTTATCATCAT 26264
Qy 111 SerLysValSerLeuThrGluLysAspClyAlaTyrThrValHisIleThrProGlyPro 130
Db 26265 ACCAAACTTTCATCGATTCGCAATTAAGCAGTCGCCGCTAGTGTGACTGGTGAAGCGGGTGT 26324
Qy 131 ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn 150
Db 26325 TTTTCACAGTCCGAGGTTTATTGAAATGTGTAAATGCT----- 26363
Qy 151 LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPhe 170
Db 26364 TTACATCAACGTTATTATTCGCGCATTAAGATTGGCAACACCG-----AGTGACAAAT 26417
Qy 171 AspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGly 190
Db 26418 CAATATGTCACGGTGGCGGAATTC-----CCATTCTATATGCTGTAGATGCTCGACCTGG 26474
Qy 191 TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
Db 26475 TTCCCATCATGCTGATGGCTCTGCACACATTCGCCCTTATATGTTGTGCTGCTGCACAGC 26534
Qy 211 AspLeuAsnValValAlaAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
Db 26535 TATTGTCATATTTCTTTTACAA----- 26555
```

```

Qy 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThr 250
Db 26556 -----GAAACGCTCTTACAAGAACTTGT---TTGAAAGAAACAGAGTC 26594
Qy 251 ProTyrAspLeuAspLeuLeuLeuAspPheGlnGlnAlaLeuGluGln----- 266
Db 26595 TTGCAAGATAATGACTTAACCTGAAAGCAAGCTTCAGACTCTTGAACAACAACATCCAGTA 26654
Qy 267 -----AsnGlyHisTyrSerGly----- 272
Db 26655 GCTGATCTCGCACCTAATGGTTACTTTGTCATCGAGCGAGTTCAGCTAATCATAGTACAA 26714
Qy 273 -----AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValPro 288
Db 26715 GCTAATGACGAACACAAATTACGCTGTGATTAACAACATATTACAGGCGCAGTTAAGTACT 26774
Qy 289 ValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAsp 308
Db 26775 ACTGGCATAAGTACTATCATCAGTATTAAA----- 26801
Qy 309 SerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyr-----AsnLeuPheAsn 326
Db 26802 -----CAGATCGCACGACACACTGTTATGCCGCTAATGATACTAATAC 26840
Qy 327 LysGlyTyr-----IleGlySerValValTrpAspMetAspLysTyrGluThr 342
Db 26841 AAAGCCTATAGCGCAGTGTATTGCGGAGACTGCTGAAGAGTTAAGCAAGAAATAAAC 26900
Qy 343 ThrLeuAlaAlaGlyLeuSerGlnProArgAsnTyr----- 354
Db 26901 TTGGCGTTTGGCTGCTATCGCTAGCTAGCTGTTTAATGACAGTCTAAAGATGGAACCCCG 26960
Qy 355 ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg-----SerThrThrGlnAsn 372
Db 26961 AAGGCGAGTATTATTTACCGCGCAGCTGCAAAATAACAGCGCTCTAACACAGCACAGAAAT 27020
Qy 373 LeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAsp 392
Db 27021 GGTGTCACTTCATGTCACCAAGT----- 27044

RESULT 37
US-08-164-292B-15
; Sequence 15, Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
```

ATTORNEY/AGENT INFORMATION:

NAME: GRACEY, NANCY J.
REGISTRATION NUMBER: 28,216
REFERENCE/DOCKET NUMBER: 29310-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 677-7000
TELEFAX: (415) 677-7522
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 5100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2..418
US-08-164-292B-15

Alignment Scores:
Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Gaps: 36

US-09-857-669-2 (1-609) x US-08-164-292B-15 (1-5100)

QY 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
Db 2131 ATTACCGCGGGGCTCTCGGACAAACGACCTCGCGTCCGCTCTCTCCCGCGTG 2190

QY 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
Db 2191 GCCTCCGATGAGCGCGGACGCTCAACCTGCTCAACCTGCTGCTATATAAGGAC 2250

QY 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe-----53
Db 2251 AACAAAGTA-----GCTGTCAAAGTAGTCCGCGGCTGCTCCGCTCCCAAT 2298

QY 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
Db 2299 ATGTCTCTCAGGTCACACAGCGGCGGCTCAGGTAACGATGACAAGGTGCTCTA 2358

QY 69 GluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLysGluGlnThr 88
Db 2359 AATACCAAGCTCCCTCTCGACACACCGCGGCGCTCTCCCTACTCTGGGTCCGAC 2418

QY 89 GlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 2419 CTCACCTTAGGTGAGGAGGAACGACTAACAGTAACACC-----2457

QY 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 2458 -----GGAGCGGGCTCCAAATAGCAATAACGCTCGCGCTTAAAGTAGTTCA 2508

QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 2509 GGTATACCGGTAGATGCTCAAAACACGCTCGCTGCATPCTCTGGGGACGGTCTA-----2562

QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
Db 2562 -----2562

QY 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----Ala 185
Db 2563 -----GAAAGCAGAGATAATAAACCTGCTTAAAGCTGGCGCGGACTTACA 2610

QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 2611 ATAACTAATCAAGCTCTTACTCTTGCT---ACCGGGAAC---GGCTTCAGGTCAACCG 2664

QY 206 AspThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGly 225
Db 2665 GAA---GGCAACTGCAGCTAAACATTAATGCGCGTCAAGGCTCAACTTTCGAACAAC 2721

QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 2722 AGCCTCGCGTG-----CAGCTGGCTCGGCGCTG---CAT 2754

QY 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu-----256
Db 2755 TTTCCCGCTGGCCAAACCAAGTAAAGCTTTATCCCGAGATGGAATAGACATCGAGAT 2814

QY 257 -----LeuLeuAspPheGlnGlnAlaLeu 264
Db 2815 AATAGGTGACTGTCCCGCTGGCGGAGGCTGAGATGCTCAACCAACCACTTGGCTA 2874

QY 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
Db 2875 GCT-----TCCGGAGACGGTTTAGAAGTCCACAGCAGCACCCCTC-----2913

QY 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
Db 2914 ---CGGTTAAAGCTCTCCCGGCTGACATTTGAAAATGGCGCGCTACGAGCAAACTA 2970

QY 302 GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
Db 2971 GGACCAAGGACTTGGCAGCAGAGGACTGTGTCGTCGCTGCTCGCAGGCTCGAGACTT 3030

QY 316 LysIleAlaTyrAspTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal 333
Db 3031 AGAGTTGCAACCGGCAAGTCCAGATCTTCAGCGGAAGAGCAGCCGCTCGGCACT---3087

QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
Db 3088 -----GATAGCAGCTCACTCTCAACATCCGCGGCGCCCTTACAA 3126

QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu 373
Db 3127 TTTTCGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3186

QY 374 GluLysArgAlaPhe-----SerGlyGlyIleTrpTyrValArgAspArgAla 389
Db 3187 ACAATGGCACTTTCCGCTCTCTATAGCCCGGAATGTGGGTAGACCAACACAGACTT 3246

QY 390 GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAla 401
Db 3247 CAGGTAACCCAGCGCTGGTGTAGCTTCCCAAGGAAACAACTTGTCCCAACCTTGC 3306

QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer 416
Db 3307 GATCCGCTGGCTATTTCGACAGACAAATTAAGTCTAGTCTCGGCTCCGCGCTGAC 3366

QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
Db 3367 GCTTCCACCGCTGCTTAAAGT-----TTAGAAACGGCTTGAATTC 3411

QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 3412 TCCAAT---CAAGCGCTGCTATAAAGCGGCGCGGCTTA---CGCTTGTAGTCTTC 3465

QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 3466 TCAAGCTTTAGAGAGCAGCTCACAGTCGGAATGGCTTAACG-----3510

QY 477 LeuGlyThrPheIleAlaArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn---494
Db 3511 CTTACCGATACTGTATCCGCGCCCAACCTAGGAGCGCGCTAGAGGTGACAGACAATAA 3570

QY 495 -----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArg 512
Db 3571 ATCATGTGTAAGCTGGCGCGAATCTTCGTTTGAACACGGCGCTGACCGC-----3624

```
QY 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
Db 3625 -----GGACGGTTAAACCTTCTCGCGCCCGAGGACCACT 3663
QY 531 -----GluArgAlaLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
Db 3664 CTCACCTGCAGACCAACCCCTCGAGCGCTCCAACTCCCATCTTCAACTGTCCTATCG--- 3720
QY 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
Db 3721 -----GAGGGTTGGTGTGCATACAAACGCCCTCTCCTCAACTGGAGACGGCATG 3774
QY 561 AlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArg 578
Db 3775 GAAGTAAATCAGCAGCGACTTACTTTAAGAGTAGGTCGGCTTGTGCAAAATGGCT 3828

RESULT 38
US-08-164-292B-17
; Sequence 17 Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE
; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 29310-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 677-7000
; TELEFAX: (415) 677-7522
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 408..1331
US-08-164-292B-17

Alignment Scores:
Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Gaps: 36

US-09-857-669-2 (1-609) x US-08-164-292B-17 (1-5100)
```

```
QY 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
Db 2131 ATTAACCCCGCGGGCTTCTGGAGACAAACGACCTCGCTCCGCTCTCTTCCACCGCTG 2190
QY 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
Db 2191 GCCTCCGATGAGCGCGCAACGCTCAGCTCAACATGCTCAGCGGCTATATATAAGGAC 2250
QY 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
Db 2251 AACAAGCTA-----GCTGTCAAGTAGGTCCCGGGCTGTCCTCGACTCCAAT 2298
QY 54 ---ProValArgIleAspThrGlnAsp-----SerGluLysLysAspMetVal 68
Db 2299 AATGCTCTCCAGGTCACACACAGCGGCTCAGCTAACCGATGACAAGGTGTCTCTA 2358
QY 69 GluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLysGluGlnThr 88
Db 2359 AATACCAAGCTCCCTCTCGACCAACAGCGCGGCTCTCCCTACTCTTGGGTCCCAAC 2418
QY 89 GlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 2419 CTCACCTTAGGTGAGGAGCAAGCTAACAGTAACACC----- 2457
QY 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 2458 -----GGAGCGGGCTCCAAATTAGCAATTAACGCTCTGGCCGTAAGAGTGTCA 2508
QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 2509 GGTATCACCGTAGATGCTCAAAAACAGCTCGCTGCATCCTCGGGGAGGCTCTA----- 2562
QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 2562 ----- 2562
QY 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly-----Ala 185
Db 2563 -----GAAAGCAGAGATAATAAACTGCTGTTAAGGCTGGCGCGGACTTACA 2610
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 2611 ATAACTAATCAAGCTCTTACTGTGCT---ACCGGGAAC---GGCTTCAAGTCAACCG 2664
QY 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 2665 GAA---GGGCAACTGCAGCTAACATTAAGTCCGGTCAGGGCTCAACTTTGCAACAAC 2721
QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 2722 AGCTCGCGCTG-----GAGCTGGCTCGGGCCTG---CAT 2754
QY 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu----- 256
Db 2755 TTTCCCGCTGGCCAAACCAAGTAAGCCTTTATCCCGAGATGGAATAGACATCCGAGAT 2814
QY 257 -----LeuLeuAspPheGlnAlaLeu 264
Db 2815 ATAGGTGACTGTCCCGCTGGCGGAGGCTGAGATGCTCAACCACTTCCCGTA 2874
QY 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheArgLeuGlnGly 284
Db 2875 GCT-----TCCGGAGACGGTTTAGAAGTCCACAGCAGCACACCTC----- 2913
QY 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
Db 2914 ---CGTTAAAGCTCTCCCGCGGCTGACATTTCAAANTGGCGCGGTACAGCAAACTA 2970
QY 302 GluThrGlyIleArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
Db 2971 GGACCAAGGACTTGGCACAGACGACTCTGGTCCGCTCGTGGTTCGCACAGGTTCGAGGACTT 3030
```



```

: ADDRESSEE: MORRISON & FOERSTER
: STREET: 345 California Street
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94104-2675
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/164.292B
: FILING DATE: 09-DEC-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: GRACEY, NANCY J.
: REGISTRATION NUMBER: 28,216
: REFERENCE/DOCKET NUMBER: 29310-20021.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 677-7000
: TELEFAX: (415) 677-7522
: TELEX: 34-0154
:
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5100 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 529..954
: US-08-164-292B-19
:
Alignment Scores:
Pred. No.: 0.19 Length: 5100
Score: 111.00 Matches: 146
Percent Similarity: 34.04% Conservative: 78
Best Local Similarity: 22.19% Mismatches: 260
Query Match: 3.50% Indels: 174
DB: 1 Gaps: 36
:
US-09-857-669-2 (1-609) x US-08-164-292B-19 (1-5100)
:
Qy 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
: ||| ||| ||||| ||| |||
Db 2131 ATTAACCCCGGGGGTCTTGGAGACAAAGACCTCGCGTCCGCTGCTTCCCAACCGCTG 2190
:
Qy 19 AlaTyAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
: ||| ||| ||| ||||| ||| ||| |||
Db 2191 GCCTCCGATGAGCGCGGCAACGTACGCTCAACATGCTACGCGGCTATATCTAAGGAC 2250
:
Qy 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
: ||| ||| ||||| ||| |||
Db 2251 AACAAAGCTA-----GCTGCAAAAGTAGGTCCCGGGTGTCCCTCGACTCCAAT 2298
:
Qy 54 --ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
: ||| ||| ||||| ||| |||
Db 2299 AATGCTCTCCAGGTCCACACAGCGGCGGCTACCGGTAAACGGATGACAGGTGCTCTTA 2358
:
Qy 69 GluGluHisLeuProLeuTleThrGlnGlnGluValLeuAspLysGluGlnThr 88
: ||| ||| ||||| ||| |||
Db 2359 AATACCAAGCTCCCTCTCGACACACAGCGGGGCTCTCCCTACTTCTGGGTCCGAGC 2418
:
Qy 89 GlyPheLeuAlaGluAlaProAsnValLysThrMetLeuArgSerLysGlyTyr 108
: ||| ||| ||||| ||| |||
Db 2419 CTCACCTTAGGTGAGGAGGACGACTAACAGTAAACACC----- 2457
:
Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyThrValHisIleThrPro 128
: ||| ||| ||| ||| |||
Db 2458 -----GGACGGCGCTCCCAATTAGCAATACCGCTTGGCGGTAAAAAGTAGGTCA 2508

```


; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1246..1707

; US-08-164-292B-21

Alignment Scores:

Pred. No.: 0.19 Length: 5100
 Score: 111.00 Matches: 146
 Percent Similarity: 34.04% Conservative: 78
 Best Local Similarity: 22.19% Mismatches: 260
 Query Match: 3.50% Indels: 174
 DB: 1 Caps: 36

US-09-857-669-2 (1-609) x US-08-164-292B-21 (1-5100)

QY 3 IleLysProThrAlaLeuLeu-----LeuProAlaLeuPhePheProHis 18
 DB 2131 ATTAACCCCGGGGCTTCTGGACAAACAGCCTCGCGTCCGTCTTCCACCGTG 2190
 QY 19 AlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLys 38
 DB 2191 GCTCGATGAGCGCGGCAAGCTCACGCTCAACATGCTGACGGGCTATATATAAGGAC 2250
 QY 39 AsnLysSerProAspThrGluSerValLysLeuLysProLysPhe----- 53
 DB 2251 ACAAGCTA-----GCTGTCAAAGTAGTCCCGGGCTGCCCTCGACTCCCAAT 2298
 QY 54 ---ProValArgIleAspThrGlnAsp-----SerGluIleLysAspMetVal 68
 DB 2299 AATGCTCTCCAGGTCACACAGCGGCGCTCACGGTAACCGATGACAAAGGTGCTCTA 2358
 QY 69 GluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThr 88
 DB 2359 AATACCAAGCTCCCTCTCGACACAGCGGGCGCTCTCCCTACTTCTGGTCCCGAC 2418
 QY 89 GlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
 DB 2419 CTCCTCTAGGTGAGGAGGACGACTAACAGTAAACACC----- 2457
 QY 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
 DB 2458 -----GGAGCGGGCTCCAAATTAGCAATAACCTCTGCGCGTAAAGTAGGTCA 2508
 QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
 DB 2509 GGTATCACGCTAGATGCTCAAAACAGCTCGCTGCATCCCTGGGGGACGGTCTA----- 2562
 QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
 DB 2562 ----- 2562
 QY 169 AspPheAspGlnAspSerTrpLeuAsnSerLysThrSerValLeuGly-----Ala 185
 DB 2563 -----GAAAGCAGAGATAATAAAACTGCTGCTAAAGCTGGCGCGGACTTACA 2610
 QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
 DB 2611 ATAACTAATCAAGCTCTTACTGTGCT---ACCGGGAAC---GGCTTCAAGGTCAACCCG 2664
 QY 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
 DB 2665 GAA---GGCAACTGCGAGCTAAACATTAATCTCGCGTCAAGGCTCAACTTTGCCAACAAAC 2721
 QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
 DB 2722 AGCTCGCCGCTG-----GAGCTGGGCTCGGGCTG---CAT 2754
 QY 246 PheGlnProGlyThr-----ProTyrAsp---LeuAspLeu----- 256
 DB 2755 TTTCCCTCTGGCCAAACCAAGTAAGCCTTTATCCCGAGATGAATAGACATCCGAGAT 2814
 QY 257 -----LeuLeuAspPheGlnGlnAlaLeu 264

DB 2815 AATAGGTGACTGTGCCCGTGGGCGCAGGCTTGAGATGCTCAACCAACCTTCCGCTA 2874
 QY 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
 DB 2875 GCT-----TCCGAGACGGTTTAGAAGTCCACAGCGACACCCCTC----- 2913
 QY 285 AspArgValProValLysValSerValThr-----GluValLysArgHisLysLeu 301
 DB 2914 ---CGGTTAAAGCTCTCCACGGCTGACATTTGAAAATGGCGCGTACAGCAAACTA 2970
 QY 302 GluThrGlyLeuArgLeuAspSer-----GluTyrGlyLeuGlyGly 315
 DB 2971 GGACCAAGACTTGGCACAGAGCACTCTGGTGGTCCGTGGTTCGCAGGTCGAGGACTT 3030
 QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsn---LysGly---TyrIleGlySerVal 333
 DB 3031 AGGTTTGAACCGCAAGTCCAGATCTTCAGCGGAAGACGACCGCCATCGGCATC--- 3087
 QY 334 ValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsn 353
 DB 3088 -----GATAGCAGCTCCTCAACATCCGGGCGCGCTTACAA 3126
 QY 354 TyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeu 373
 DB 3127 TTTTCGGACCGCTTCTGCTAGTTTGGACGAGCTGGTCCGATTCTTACAAACG 3186
 QY 374 GluLysArgAlaPhe-----SerGlyGlyIleTrpTyrValArgAspArgAla 389
 DB 3187 AACATGGCATTTCGGTCTCTCTATAGCCCGGAATGTGGGTAGACCAAAACAGACTT 3246
 QY 390 GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAla 401
 DB 3247 CAGGTAAACCCAGGCGGTGGTTTACTCTTCCAGGAAACACCTTGTCCAAACCTTGG 3306
 QY 402 GluGlyArgLysIleProGlySerAspIleAspLeuGly-----AsnSer 416
 DB 3307 GATCCGCTGGCTATTTCGACAGACGAAATATTAGTCTAGTCTCGGTCGCGCTGACCCAA 3366
 QY 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisPro 436
 DB 3367 GCTTCCACCGCTGACTTTAAGT-----TTAGAAACGGGCTTGAATTC 3411
 QY 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
 DB 3412 TCCAAT---CAAGCGCTTGTCTATAAAAGCGGGCGGGCTTA---CGCTTGAAGTCTTC 3465
 QY 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
 DB 3466 TCACAAAGCTTTAGACAGAGCGCTCACAGTCGGAATGGCTTAACG----- 3510
 QY 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrVal---AlaArgAspAsn--- 494
 DB 3511 CTTACCGATAGTGTGATCGCCCAACCTAGGGACGCGCTAGAGGTGACAGACAATAAA 3570
 QY 495 -----AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArg 512
 DB 3571 ATCATTGTTAAGTGGCGCGAATCTTCGTTTGAACAGGAGCGGTAAACCGC----- 3624
 QY 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuPro----- 530
 DB 3625 -----GGCACCGTTAACGCTTCTGCGCGCGAGGACCAACCACT 3663
 QY 531 -----GluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
 DB 3664 CTCACCTGCAGACACCCCTCCGAGCCTCCCACTCCCACTTTCACCTGTCCTATCG--- 3720
 QY 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
 DB 3721 -----GAGGCGTGGTTCTGCATAACAAACCGCTTGTCTTCCAACTGGGAGACGGCATG 3774
 QY 561 AlaAlaAsnPhelysArgMetLysLysHisGlySerGlyLeuGlyValArg 578

Db 3775 GAAGTAAATCAGCACGGACTTACTTTTAGAGTAGGCTCGGGTTTGCAAAATGCGT 3828

Search completed: November 9, 2002, 02:58:46
Job time : 226 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 02:46:30 Search time 61 Seconds

(without alignments)

3542.038 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 1713

Sequence: 1 MIKPTALLPALEFFPHAY.....IAYGSDKKIRWHISLGTRF 609

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Ygapext 0.5
Ygapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-DB=published_applications_NA -QPM=fastap -SUFFIX=us09857669/runat_05112002_110010_7843/app_query.fasta_1.775
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09857669.ecgn_1.1.21.ernat_05112002_110010_7843
-NCPU=6 -ICPU=3 -NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	4.6	10419	9 US-10-027-806-3	Sequence 3, Appli
2	146	4.6	42432	9 US-10-027-806-2	Sequence 2, Appli
3	129.5	4.1	1398	9 US-09-938-842A-2084	Sequence 2084, Ap
4	124.5	3.9	2448	10 US-09-815-242-4051	Sequence 4051, Ap

5	122	3.8	6026	9 US-09-995-749A-1	Sequence 1, Appli
6	121.5	3.8	640681	10 US-09-790-988-1	Sequence 1, Appli
7	116	3.7	2950	10 US-09-886-468-6	Sequence 6, Appli
8	112	3.5	2439	10 US-09-974-300-472	Sequence 472, App
9	112	3.5	3213	10 US-09-974-300-475	Sequence 475, App
10	112	3.5	3286	10 US-09-332-226-1	Sequence 1, Appli
11	111	3.5	34446	10 US-09-871-212-1	Sequence 1, Appli
12	108.5	3.4	1969	10 US-09-864-761-4449	Sequence 4449, Ap
13	108.5	3.4	2004	10 US-09-864-761-21197	Sequence 21197, A
14	108.5	3.4	4896	10 US-09-740-274-3	Sequence 3, Appli
15	108.5	3.4	9477	10 US-09-815-242-8513	Sequence 8513, Ap
16	108	3.4	4065	12 US-10-044-090-98	Sequence 98, Appl
17	107.5	3.4	14800	10 US-09-954-456-1601	Sequence 1601, Ap
18	106.5	3.4	2140	10 US-09-974-300-2290	Sequence 2290, Ap
19	106.5	3.4	2140	10 US-09-974-300-2620	Sequence 2620, Ap
20	106.5	3.4	3537	10 US-09-332-226-3	Sequence 3, Appli
21	104.5	3.3	1149	10 US-09-912-020-139	Sequence 139, App
22	104	3.3	2055	10 US-09-974-300-2362	Sequence 2362, Ap
23	103.5	3.3	4557	10 US-09-974-300-2632	Sequence 2632, Ap
24	103.5	3.3	5480	10 US-09-070-927A-132	Sequence 132, App
25	103	3.2	1164	10 US-09-974-300-28	Sequence 28, Appl
26	102.5	3.2	1389	9 US-09-938-842A-2606	Sequence 2606, Ap
27	102.5	3.2	22671	10 US-09-956-004-14	Sequence 14, Appl
28	101.5	3.2	15231	10 US-09-917-800A-1505	Sequence 1505, Ap
29	101	3.2	2949	10 US-09-841-132-170	Sequence 170, App
30	101	3.2	3021	10 US-09-841-132-182	Sequence 182, App
31	101	3.2	7107	10 US-09-070-927A-527	Sequence 527, App
32	100.5	3.2	2820	10 US-09-815-242-6700	Sequence 6700, Ap
33	100.5	3.2	13884	10 US-09-070-927A-341	Sequence 341, Appl
34	100.5	3.2	32768	10 US-09-070-927A-17	Sequence 17, Appl
35	100	3.2	6258	10 US-09-815-242-4584	Sequence 4584, Ap
36	100	3.2	17388	10 US-09-815-242-8512	Sequence 8512, Ap
37	100	3.2	35100	10 US-09-782-378A-26	Sequence 26, Appl
38	99.5	3.1	2085	10 US-09-974-300-7800	Sequence 7800, Ap
39	99.5	3.1	3078	9 US-09-938-842A-1039	Sequence 1039, Ap
40	99.5	3.1	3412	10 US-09-897-056-6	Sequence 6, Appli
41	99.5	3.1	6077	10 US-09-815-242-4648	Sequence 4648, Ap
42	99.5	3.1	6114	9 US-09-938-842A-2700	Sequence 2700, Ap
43	99	3.1	1243	10 US-09-070-927A-504	Sequence 504, App
44	99	3.1	1800	8 US-08-973-028-3	Sequence 3, Appli
45	99	3.1	2255	10 US-09-747-521-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-027-806-3
; Sequence 3, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 10419
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(10419)
US-10-027-806-3

Alignment Scores:
Pred. No.: 9, 28e-06 Length: 10419
Score: 146.00 Matches: 128

Percent Similarity: 31.17% Conservative: 59
Best Local Similarity: 21.33% Mismatches: 213
Query Match: 4.60% Indels: 200
DB: 9 Gaps: 31

US-09-857-669-2 (1-609) x US-10-027-806-3 (1-10419)

QY 138 GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
DB 6259 GGCAGCAGCGCCTGAATGCGGTAATATCGCAGAGGGCTCGACACCTGTACAGCTTT 6318

QY 158 AlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
DB 6319 CTACTGGAC-----ATCCCGCTATGGGGCGGAATTGATATTCACAGGGCTTGAG--- 6366

QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
DB 6367 -----CTTCCGCTGGGGGTTCCGACG-----GGAATTCGATTCG----- 6405

QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
DB 6406 -----GAC 6408

QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
DB 6409 AACGGGCGCCAGTTGTACATGGCGGCTTCGT-----GACTCTCAATCTCGCCAGGC 6462

QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu--- 256
DB 6463 ACCCTGCGTGGCGCCTGCAGCGCTATGAGCTTGGCATACCATATGACCTGGCTTCGGCT 6522

QY 257 -----LeuLeuAspPheGln-----GlnAlaLeuGluLysAsn 267
DB 6523 GTATTGGCGAGTCCCTGGGNAATATTCGATTTCTCCCTCAACGGCATGCGGGCCAA 6582

QY 268 GlyHisTyrSerGlyValaSerValGlnAlaAsp-----PheAspArgLeuGln 283
DB 6583 GGCAGCTGGCAGGATTACATGTGCGCGCCGATGGAAGCATCTGTTCAGGCGCGGAAT 6642

QY 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
DB 6643 GCCGAACACCGTAATCAGCTAT-----GACATGACAGCCATGATTTGGATACA 6693

QY 304 -----GlyIleArgLeu 307
DB 6694 TTATCATTTACGGGAATCATTTCAACACAGATGTCGGACAGTCGACACCCCAACATAAG--- 6750

QY 308 AspSerGluTyrGlyLeuGlyGlyLysIleAlaTyr-----Asp 320
DB 6751 GACATGGACATATCCCGGACGGCATGTTCTCTACCTGCTTCAAGCGCATGTTCTGAC 6810

QY 321 TyrTyrAsnLeuPheAsnLys-----GlyTyrIleGlySerValVal 334
DB 6811 ARGTAACACCTTACAGATAGTATTCGCTTGATGCCCGGCATATCGGGTACCCTG--- 6867

QY 335 TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr 354
DB 6868 --GATTTTGAACCG--GAGGATGTAATACCCAGGGGATTTTCTCTCAGGGGATGCG 6921

QY 355 ArgGlyAsnTyrTrpThrSer----- 361
DB 6922 ACGACTCTGTTATGACAGGGGAGAGAGGTGGACACCATTCACGAATATGCAATGAATGAA 6981

QY 361 ----- 361

DB 6982 CCATGGGACATACGAATGCCATACTTGCAGGCTCCCTCTCCATAAGCGCAGTGAATGGT 7041

QY 362 -----AsnValSerTyrAsnArgSerThrThrClnAsnLeuLysArg 376
DB 7042 GCACCGCGGGGCTGGATATATCGAGGATGGCAACACTGCACATACTATGCGCGGGCT 7101

QY 377 AlaPheSerGlyIleTyrTrpValArgAspArgAlaGlyIle----- 391
DB 391

DB 7102 GACTTTTGACACGGG-----CCCGCATCCCTGGTAAACACCATATTG 7143

QY 392 -----AspAlaArgLeuGlyAlaGluPheLeuAlaGluGly 403
DB 7144 CCAGCCCAATATTCCTGCTGACGGATGCGCGCTTTGCA---TACCCCTGGAGAG 7200

QY 404 ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
DB 7201 GAGGGTGCACCGGG---GATCTTGCAATTCCTCGATGACGGCATGGCATGTCGTGGC 7257

QY 424 SerTrpLysArgGlnLeuLeuAsn---AsnValLeuHisPro-----GluAsnGly 439
DB 7258 GCGCTAAACAACCACTTTAAAGACAGTACAACTGCTGCTGCGTATGACACTGAAATGCA 7317

QY 440 ---HisTyrLeuAspGlyLysIle-----GlyThrThrLeuGlyThrPhe 453
DB 7318 GAACATTTCACTCGACGGATCTGCTGACTGCGGACAGGGCCCGGCTTGTATT 7377

QY 454 LeuSerSerThrAlaLeuIleLeuThrSerAlaArgAlaGlyTyr-----PhePhe 470
DB 7378 TCAGATGAGAACGACTTTTTCAGCACAGCGCCAGGGCCCAATTTGTGCGCCAGTTTACG 7437

QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrVal 490
DB 7438 ACAAAACCGCCCTACGACGGCATCCACAATAACACTGACTGACACAGGACTGTACAAGTG 7497

QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSer 510
DB 7498 AGCGTGGAGCGC-----CTGCGCTCGGCATACGGTTTACCCCGCAGGCAATGAAGATG 7551

QY 511 ---ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeu 529
DB 7552 TTCAATATCGGCGGACGAGAGCGCCATGATA----- 7581

QY 530 ProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeu 549
DB 7582 -----TACAGTATTCCTGCTGCGCTCCCGCTGATGACACA 7614

QY 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
DB 7615 TCCGGGGCGGTGAGGATGAGATAGTCGACGGGCTCTTTAGAAATCAGGTTG 7674

QY 570 Lys----- 570

DB 7675 TCCGTGCGGTTGAACGAGCCAGTCCTTCGCGCTTTGACTTTTCGGAGGACGGAATGGAG 7734

QY 571 -----HisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSer 587
DB 7735 CTGTACGTGACGGGTGCGGCTTGTTCACAGGTATTTCCTGCCA---TCGCCATACGGC 7791

QY 588 Phe---AspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGly 606
DB 7792 CTCGAAGATGACGCTACGGGGGCGAGCTTCCACAGCTTCAGGAGAGACGCGCGCTGGGA 7851

RESULT 2
US-10-027-806-2
; Sequence 2, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schlepman, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DORP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA

ORGANISM: Cenarchaeum symbiosum

```

FEATURE:
NAME/KEY: CDS
LOCATION: (3)...(10421)
NAME/KEY: CDS
LOCATION: (10625)...(11434)
NAME/KEY: CDS
LOCATION: (11478)...(13046)
NAME/KEY: CDS
LOCATION: (13046)...(14620)
NAME/KEY: CDS
LOCATION: (23558)...(24862)
NAME/KEY: CDS
LOCATION: (24913)...(25728)
NAME/KEY: CDS
LOCATION: (26504)...(26881)
NAME/KEY: CDS
LOCATION: (29655)...(30491)
NAME/KEY: CDS
LOCATION: (34559)...(36067)
NAME/KEY: CDS
LOCATION: (37002)...(37403)
NAME/KEY: CDS
LOCATION: (37404)...(38282)
NAME/KEY: CDS
LOCATION: (39454)...(40572)
US-10-027-806-2

Alignment Scores:
Pred. No.: 8,07e-05
Score: 146.00
Percent Similarity: 31.17%
Best Local Similarity: 21.33%
Query Match: 4.60%
Dbs: 9
Gaps: 31

US-09-857-669-2 (1-609) x US-10-027-806-2 (1-42432)

QY 138 GlyValAlaLeuLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsn 157
Db 6261 GCGAGCAGCGCCTGAATGCGTAATATCGCAGAGGGCTCGACACCTGTACAGCTTT 6320

QY 158 AlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAsnSerTrpGluAsn 177
Db 6321 GTACTGGAC-----ATCCCGTATGGGGCGGAATTGGATATTGACAGGCTTGAG--- 6368

QY 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 6369 -----CTTCGCGTGGGGGTTCCGAGC-----GGATTGAGTTCTCG----- 6407

QY 198 AsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValAsp 217
Db 6408 -----GAC 6410

QY 218 SerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGlu 237
Db 6411 AACGGGCGCCAGTTGTACATTGGCGGCTTTCGT-----GACTCTCAATCCTCGCCAGGC 6464

QY 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu--- 256
Db 6465 ACCCTGCTCGGGGCGCTGACCGCTATGAGCTTGGCATACCATATGACCTGGCTTCGGCT 6524

QY 257 -----LeuLeuAspPheGln-----GlnAlaLeuGluGlnAsn 267
Db 6525 GTATTTCGCGAGTCCCTGGGAATATTTCGATTTTCCTCCCTTCACGGCATGGCGCCCAAT 6584

QY 268 GlyHISLysThrSerGlyAlaSerValGlnAlaAsp-----PheAspArgLeuGln 283
Db 6585 GCGAGTTGGCAGGATTACATGTGCGCCCGCATGGAAGCATCCTCTGTTCAGGGCCGGAAT 6644

QY 284 GlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThr 303
Db 6645 GCGAAAGAACCGTAATCAGCTAT-----GACATGGACAGCATGATTGGATACA 6695

```

```

QY 304 -----GlyIleArgLeu 307
Db 6696 TTATCATTCAGGGAATCATTTCAAACCCAGATGTGGACAGTCGACACCCCAACATAAGG--- 6752

QY 308 AspSerGluTyrGlyLeuGlyLysIleAlaTyr-----Asp 320
Db 6753 GACATGGACATATCCCGGACGGCATGTTCCCTACCTGCTTCAAGGCGATGTTCTTGGAC 6812

QY 321 TyrTyrAsnLeuPheAsnLys-----GlyTyrIleGlySerValVal 334
Db 6813 ATGTACAACTTACAGATAGTATTTCGCTTGATGTCGCCGGCATATCGGGGTACCTG--- 6869

QY 335 TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr 354
Db 6870 ---GATTTGGAAACGG---GAGGATGTAATACCCAGGGGATTTTCATTCTCACGGGATGGC 6923

QY 355 ArgGlyAsnTyrTrpThrSer----- 361
Db 6924 ACGAGTCTGTATTATGACAGAGCGAAGACGTGGACCAATTCACGAATATGCATTTGAATGAA 6983

QY 361 ----- 361
Db 6984 CCATGGGACATACGCAATGCCATACTTGCAGGCTCCCTGTCCATAGCGCAGTGAATGGT 7043

QY 362 -----AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 7044 GCACCGCGGGGCTGGATATATCGAGGATGGCACAACATGCACATACTATGCGCGGCGGT 7103

QY 377 AlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIle----- 391
Db 7104 GACITTTGACACGGG-----CCCGCATCCCTGCTAAACCCACACATATTTG 7145

QY 392 -----AspAlaArgLeuGlyAlaGluPheLeuAlaGluGly 403
Db 7146 CCAGGCCAATATTCCTGCTGACGGATGCGCGCGCTTTTGCA---TACCCGCTGGAGAG 7202

QY 404 ArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla 423
Db 7203 GAGGGTGACCGGGG---GATCTTGCAATTCCTCGATGACGGCATGTCGTGGCG 7259

QY 424 SerTrpLysArgGlnLeuLeuAsn---AsnValLeuHisPro-----GluAsnGly 439
Db 7260 GCGGTAAACAACCATTTTAAAGACAGTACAACTGCTGTCGCGCTGTACATGACATAAATGCA 7319

QY 440 ---HisTyrLeuAspGlyLysIle-----GlyThrThrLeuGlyThrPhe 453
Db 7320 GAACATTTTCATCTCGACGGATCTGCTGACTGCGGACAGGGGCCACGGGCTTGTATT 7379

QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyr-----PhePhe 470
Db 7380 TCAGATGAAACGACTTTTTCAGCACAGCGCGCCAGGCCCAATTTGTGCGCCAGTTTACG 7439

QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleArgGlyGlnAlaGlyTyrThrVal 490
Db 7440 ACAAAACCCCGCTACGACGGCATCCACAATAACACTGATGACACACGACTGTACAAGTG 7499

QY 491 AlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyAlaSerSer 510
Db 7500 AGCGTGGACGCGC-----CTGCGCTCGGCATACGTTTACCCCGACGGCATGAAGATG 7553

QY 511 ---ValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeu 529
Db 7554 TTCATATCGGGCCAGGAGACGGCCATGATA----- 7583

QY 530 ProGluArgAlaLeuLeuValGlySerLeuGlyTyrGlnLeuProPheThrArgThrLeu 549
Db 7584 -----TACCAGTATTTCCTGCGCTCCCGGTATGACACA 7616

QY 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPhelysArgMetLysLeu 569
Db 7617 TCCGGGCGGTCAGGACAGGGTTGAGATAGTCGACAGGGCTCTTTAGAAATACGAGTTTG 7676

```

QY 570 Lys----- 570
Db 7677 TCCGTCGGTGTGAACGAGCCAGTCCTCCGGCTTTGACTTTTCGGAGAGCGAATGGAG 7736
QY 571 -----HISGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSer 587
Db 7737 CTGTACGTGACGGGTCGGGCTTGTTCACAGGTATTTCCTGCCA--TCGCCATACGCC 7793
QY 588 Phe---AspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHisIleSerLeuGly 606
Db 7794 CTCGAAGATGACGCTACGGGGGACGCTTCCACACGTTCCAGGAGAGCAGCCGCTGGGA 7853
RESULT 3
US-09-938-842A-2084
: Sequence 2084, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: FILE REFERENCE: SRIPI300-3
: CURRENT APPLICATION NUMBER: US/09/938,842A
: PRIOR FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300,111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 2084
: LENGTH: 1398
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-938-842A-2084
Alignment Scores:
Pred No.: 3,16e-05 Length: 1398
Score: 129.50 Matches: 121
Percent Similarity: 35.50% Conservative: 54
Best local Similarity: 24.54% Mismatches: 198
Query Match: 4.08% Indels: 121
DB: 9 Gaps: 26
US-09-857-669-2 (1-609) x US-09-938-842A-2084 (1-1398)
QY 96 ProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeu 115
Db 103 CCTTCGAATTCGAAACT-----CGTAGCTTAAGAACCATCTCTGTACGAGCTGGCTTA 156
QY 116 ThrGluLysAspGlyAlaTyrThrValHis---IleThrProGlyProArgThrLysIle 134
Db 157 ATCGAGCAGATGCTGGGAAACTTGTGGATCTTGTGTACCGGAACCGAGAGCGGCGAG 216
QY 135 AlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspClyAsnLeuAlaGluTyr 154
Db 217 AGAAACACGAGCGGGGATTTGCCGAGATGAGATTGACGGCGATTGATTGGCAATGG 276
QY 155 TyrArgAsnAlaLeuGluAsnTrpGlnProVal-----GlySerAsp 169
Db 277 ATGCATGTGTTGAGTGAAGGTTGGGCTAGTCTCTCGTGGGTTTATGAGGGAATCTGAG 336
QY 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
Db 337 TTCTCCAAACTTTCATTTCAATTTGTTGAATCTCGATGATGGGTCTGTTGTTAATATG 396
QY 190 GlyTyrProLeuAla---LysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla 208
Db 397 TCTGTGCTATTGTTCTTGCGATTGATGATCAACAAAGGCCCTAATCGGTGAATCTAAA 456

QY 209 ThrValAspLeuAsnValValAspSer---GlyArgProIleAla---PheGlyAsp 226
Db 457 CGTGTCTCCCTTT-----GTTGATTTGATGATAAATCAATCGTCTTCTCAATGAT 507
QY 227 PheGluIleThrGlyThrGlnArgTyrPro-----GluGlnIleValSerGlyLeuAla 244
Db 508 ATTGAGATT-----TATAACATCCGAAGAGAGGAATACCGAGAAGTTCGGGT 558
QY 245 ArgPheGlnProGlyThrProTyr----- 252
Db 559 ACGACTGACCGGGTTTGCCTTATGTAGAAAGAGCGGATAACCAATGCTGGAGACTGGCTC 618
QY 253 -----AspLeuAspLeuLeu----- 257
Db 619 ATTTGGGGTGATCTTGGAGTTTGGAACTTGTAACTACATCATGGCTTTCATCGTTTC 678
QY 258 -----LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAla 273
Db 679 AGGCTTTCCCGTTTGAACCTGCGTAAGGAGCTAGAGAACGTTGGTCCGGATCGGCTCTT 738
QY 274 SerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProVal----- 289
Db 739 GGTTCACGCTT-----AGGAACCCAGTTCATAATGACAT 774
QY 290 LysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg-----Leu 307
Db 775 GCTCTCTTATGACTGATCTACTCTAGGAGACTTCTTGGATGGGTATATAAACCCCTATC 834
QY 308 AspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys 327
Db 835 CTTTTCCTTCATCCATGGGAGG-----TTTACAAA 867
QY 328 GlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGly 347
Db 868 GCTGATGATGTTCTCTAAGCTGGCGAATGAACACGACGACGAGGCTGATAGATCGT 927
QY 348 IleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArg 367
Db 928 GTTCTTCATCCAGAGACTACTGTGTTTCCATATCCCATCTCCCAATGCTCTATCGTGT 987
QY 368 SerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleThrTyrValArgasp 387
Db 988 CCAACCCGAAGTCCAA-----TGGCACCCGAAGGCT 1017
QY 388 ArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIlePro 407
Db 1018 AGG-----ATCAATGCT-----GGTGCTAACTTCTACATTGCTCGTAGGATCCGCT 1065
QY 408 GlySerAsp-IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAla----- 423
Db 1066 GGATGGGTCTATCCCGTGGAGAACGCTGATCTGTACGATCCTGATCACGGGGAAGAGTT 1125
QY 424 ----SerTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLe 442
Db 1126 CTAAGCATGG-----CTCCT 1140
QY 442 uAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgTh 462
Db 1141 GGACTCGCAACGACTCAACA-----TTCCTTCCTTCA-----GGTCTGCTCGGTAC 1185
QY 462 rSerAlaArgAlaGly-----TyrPhePheThrProGluAsnLysLysLeuGlyThrPh 480
Db 1186 GAT-AAGACACAAGGAAGATGGCTTTCTTGTATCCCTCAAGGGCTTCAGGACTCTTGT 1244
QY 480 eIleIleArgGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerG1 500
Db 1245 C---ATTTCCTGGCACTAAGATGAGCATTTGGCAAGACAGACAGAGAACCCACAGATGG 1301
QY 500 yLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleG1 520
Db 1302 ATTTATGTCCCTGGAGGCTGGAAGTCCCTTGTGGATTACTAC-----GACAGTTTGAC 1355
QY 520 yLeuAlaGlyProAsnGlySerValLeuProGluArg 532


```

Db 1824 CGAACCCGACGAGCCAGCAACTACGAGATCGGCATCAAGGGTGAATATCTCCAGCGAGC 1883
Qy 445 sGleGlyThrLeuGlyThrPheLeuSerThrAlaLeuIleArgThrSerAlaAr 465
Db 1884 ACTCAATACCACTCTCGCTACTTTC---GAGATTTCATGAGGAAACCGCGGAGGAAGA 1940
Qy 465 gAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleArgGly-- 484
Db 1941 CCCCCTCTACACACGACCAACCGCAACCCGCCATC---ACCTAGCGCTACAGGGCAT 1997
Qy 484 ----- 484
Db 1998 CAAGCCCAAGACCAAGCGCTACGAGCGGAAATCTGGGTGAACCTGGCCCGCAGGTGCA 2057
Qy 485 ----GlnAlaGlyTyrThr-----ValAlaArgAspAsnAlaAspValProSerGlyLe 501
Db 2058 GGTTCAGGCTGGCTATACCCACAGATCATTCGCGGAGC-----AGTGGCAA 2105
Qy 501 uMetPheArgSerGlyAlaSer-----SerValAr 512
Db 2106 GAAGGTTTCCACCTG-GGAGGCTCAGGACCAAGTTGAGCCCTACACCAAGTTCA 2164
Qy 512 gGlyTyrGluLeuAspSerIleGly-----LeuAlaGly----- 523
Db 2165 AGCGCGCCTGGACAGCTCAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2224
Qy 524 -----ProAsnGlySerValLeuProGluArgAlaLeuLe 535
Db 2225 AGATGGTCTACAACAATCCCGCGAGCGCTGGGAAAGTTCTCCAGGAAGA---CTACT 2281
Qy 535 uValGlySerLeuGluTyrGlnLeuPro-PheThrArgThrLeuSerGlyAlaVal 553
Db 2282 GCGTGGTGGACCTGATGCGCGCTACCAGATCACCGNCAAGCTGTCCAGCCAGCGTC 2337
```

RESULT 5

```

US-09-995-749A-1
; Sequence 1, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043368-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 6026
; TYPE: DNA
; ORGANISM: Lactobacillus reuteri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (161)..(5503)
US-09-995-749A-1
```

```

Alignment Scores:
Pred. No.: 0.00213 Length: 6026
Score: 122.00 Matches: 128
Percent Similarity: 30.17% Conservative: 85
Best Local Similarity: 18.13% Mismatches: 231
Query Match: 3.84% Indels: 262
DB: 9 Gaps: 34
```

US-09-857-669-2 (1-609) x US-09-995-749A-1 (1-6026)

```

Qy 35 AlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysProLysPhePro 54
Db 3485 TCATTAGTGTGTCGTAATAAATATCCACGAAATGTTGTAATT---CCTAATTACAGT 3541
Qy 55 ValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuProLeu 74
Db 3542 TTC---GTTCCGGCTCAGCAATAAATCTCAAGATCNMAATTCAAAATGCTATT----- 3592
Qy 75 IleThrGlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGlu 94
Db 3593 -----CGTGATGTAAACAGGCAAGATTTACCACTTTTCTTCTTGAAGAT 3637
Qy 95 AlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 3638 GAGCAAAAGGGTATTGATGCGCTACATTCAAGATCAAAATTCACACAGTGAAGAAATATAAC 3697
Qy 115 LeuThrGluLysAspGlyAlaTyrThrValHisIlePheProGly-----ProArg 131
Db 3698 CTTTATATATTCGCGCTTCATACCGCAATCTTTTAACTAACANGATACATTCACCT 3757
Qy 132 ThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly----- 149
Db 3758 -----GTATAGTGTGCTGACTTGATGTACTGATGCTGCGCAA 3793
Qy 150 -----AsnLeuAlaGluTyrTyr----- 155
Db 3794 TACATGGAACATCAACACAGTTACTATGATACTTTAAGAACCTGCTTAATACAGAGTT 3853
Qy 156 ----- 158
Db 3854 AAGTATGTTCCGGTGGCCAATCAATGCAAAACAATAGCGTTGGCGGCAATAATAACATT 3913
Qy 159 LeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSer---TrpGluAsn 177
Db 3914 TTAAGTAGTGTGTTGTTATGTTAAAGTGGCATACACCTACTACTACTGCTACTACTGAA 3973
Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGly 197
Db 3974 ACCAGAACAAGGTATTGGGTTGTTGTA---AGTAATACGCCAAATCTAAGAGTAGGT 4030
Qy 198 -----AsnThrArgAla 201
Db 4031 GTCAACGATAAAGTAGTCTTCATATGGGACCTGCGGCAAGAACCAACAAATATCGGCA 4090
Qy 202 AlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGlyArgPro 221
Db 4091 GCGGTGTTGACCACTGATGAGTGC---ATTAAATTATCTTCTGTCAAGGGGACCG 4147
Qy 222 IleAlaPhe-----GlyAspPheGluIle----- 229
Db 4148 GTTGCAATGACTGACGAGACGGTATCTATCTATCTAGTACATAACCTAGTGTGTAT 4207
Qy 230 -----ThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeu 243
Db 4208 GGTAAAGAAGACAGATACAGCTGTTCAAGTTATGCTAAACCTGATCTTTCAGGATAT 4267
Qy 244 -----AlaArgPheGlnProGly 249
Db 4268 CTTGCTGTATGGGTACCAAGTTGGAGCAAGTATNCCAGATGCTCGAAGTCTCCATCT 4327
Qy 250 ThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlnAsnGlyHis 269
Db 4328 ACT-----GAAAAGAAATATGTGT 4345
Qy 270 TyrSerGlyAlaSerValGlnAlaAspPheAspArg-----LeuGlnGlyAspArg 286
Db 4346 AACTCTGCATACAGAACAAATGCTGCTTTGATTCAANTGTTATTTTGAAGCCTTTCT 4405
Qy 287 ValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg 306
Db 4406 AACTTTGTCTATACACCAACAAAGGAAGTGAAGCTGCTAATGTTCCAAATGCCCAAAAT 4465
Qy 307 LeuAspSerGluTyrGlyLeuGly-----GlyLysIleAlaTyrAspTyrAsn 323
```

```

Db 4466 GCTGATTCCTTGGTTCATAGGTTTACTTCTTCGAGATGGCCCAATATAATCA 4525
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 324 LeuPheAsnLysGlyTyrIleGlySerValValTrpAsp-----MetAspLys 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4526 AGTAAAGATCCACATTCCTAGATTCACAAATGATTAACGGATATGCCCTTACTGATCGT 4585
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 340 TyrGluThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyr--- 358
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4586 TATGAT-----CTTGGAGTGTGAGCCTAATAAGTACGGACAGATGAAGAT 4633
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 359 -----TrpThr 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4634 CTACGTATGCCATTCAGCGCTCCATAAGCTGGCTTACAAAGTAAATGGCGGATGGGTT 4693
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 SerAsnValSerTyrAsnArg-----SerThrThrGlnAsnLeuGluLys 375
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4694 CCTGACCAATCTATAACCTCTCTCGAAAGAAGTGTGTACAGTCACTCGAGTAGATGAT 4753
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 376 ArgAlaPheSerGlyGlyIleTrp-----TyrValArg 386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4754 CGT-----GGTAATGTATGGAAGATGCTATCATTAATAATAATCTGTGTGTGT 4804
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 387 AspArgAlaGly-----IleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4805 AATACTATTGGTGGCGAATACCAGAGAAGTATGGTGGAGCATTCCTCGATAAGTTA 4864
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 404 ArgIlyIle-----ProGlySerAspIleAsp 412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4865 CAAAACCTTTATCTCGAAATCTTCACAAAGAAGCAAGTTTCAACTGGTGTGCTATTGAT 4924
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn--- 431
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4925 -----CCTTCACAAAGATAACTGATGTGCAGCAAAATACTTAAATGGA 4969
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 432 ---AsnValLeuHisProGluAsnGlyHisTyrLeu-----AspGly----- 444
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4970 ACAACATCTCTCCATCGTGTCTCTGTTATGTACTAAAGCTAGTGGTGGTCAATACTAC 5029
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 445 LysIleGlyThrLeuGlyThrPheLeuSer----- 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5030 AACTAGTACTACTACAAAGCAATCTTGCCAAATTCATTAATTAACGTGGTCAAAAGAACAA 5089
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 455 ----- 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5090 CGAAATGAAGCTTTGTTAAGGCTAATGATGGAATTAATCTTCTATCACTTAGCAGGT 5149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 456 -----SerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro 472
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5150 AATATGGTTAAGATACCTTTATTGAAGATAGTGTGGCAACTGCTACTCTTTCACCAA 5209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 473 GluAsnLysLysLeu-----Gly 478
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5210 GATGTAAGATGGTTGAAATAAACAATTTCTGTTGATGTTGATCTTATGGTGAAGAAAGGT 5269
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAsnAlaAlaValPro 498
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5270 ACTTACTTCTTC-----TTGAAG 5287
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 499 SerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGly---TyrGluLeuAsp 517
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5288 AATGGTGTATCATCTCCGTGGGGATTAAGTGCAAACACTGACAATGGTACTTATTACTTTGAT 5347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 518 SerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGly 537
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5348 AATTAT-----GGAAGATGCTAGCTAATCAAACTATTAAATCCAGGT 5389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 538 SerLeuGluTyrGlnLeu 543
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5390 GCCATGATTTATACCTTA 5407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
US-09-790-988-1

```

```

; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Alignment Scores:
Pred. No.: 3.22 Length: 640681
Score: 121.50 Matches: 58
Percent Similarity: 37.62% Conservative: 62
Best Local Similarity: 18.18% Mismatches: 154
Query Match: 3.83% Indels: 45
DB: 10 Gaps: 9

US-09-857-669-2 (1-609) x US-09-790-988-1 (1-640681)
Qy 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPhe-----SerSer 111
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261638 GAACATTTTAAAAAATTTTATTTGAGTAATGGGTATTTATTTTAAATGTAACACTAAA 261697
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 112 LysValSerLeuThrGlu-----LysAspGlyAlaTyrThrValHisIleThrProGly 129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261698 AAAGTAGATATTTTTCGAAGATAAAAAACAGGTGGATATCACAATAGGTGTTTCTGAAGGA 261757
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 130 ProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGly 149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261758 AGAATAACAATAATTCCAATTTTGTGTC-----AATGGA 261793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 150 AsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAsp 169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261794 AATTTTATTCCTATCAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 261853
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 170 PheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261854 TATAATCGACACAAATGATATATATAGTAATAAATAACAAGATTTTATCTGAATAT 261913
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 190 GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp----- 206
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261914 GGATATGTT-----AATACTAAAGTTATAGTTGACCGCAAAATTTGATCAT 261958
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 207 ---ThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGly 225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261959 AAAAAACAAACAAATAGCCCTTAATTTTAAATATGATATGAAAAACGCTTATTTGTAAAA 262018
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262019 AGAATTTATTTTACAGGGAATTAACCAAGATAGAGTTTTAGTCGTGCTGCTAAAAATTA 262078
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 246 PheGlnProGlyThrProTyrAspLeuAspLeuAspPheGlnGlnAlaLeuGlu 265
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262079 CAATGGGAGGTAAATATTTTAAATACAAATTTAGTAGAATAGGTAAGTAACATATTAGAA 262138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 266 GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAsp 285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262139 AAAAAAATATTTTGTAGTGTAAAAAATCATTCAAAAAATTAATTTCTTATGATTTCTAAT 262198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 286 ArgValProValLysValSerValThrGluValLysArgHisLysLysLeuThrGlyIle 305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 262199 CAAATTGATATTACCTATCAAGTTAAAGACAGACTACTGGTCTATATAAAATTTTGGTTTA 262258
QY 306 ArgLeuAspSerGluTyrGlyLeuGlyGlyLeuAlaTyrAspTyrTyrAsnLeuPhe 325
Db 262259 GGATATGGAGTAGTAGTGACAGTTTAACTCTTGCTTTTCTCAAGATAATATATTT 262318
QY 326 AsnLysGlyTyrIleGlySerValIleTyrAspMetAspLys-----TyrGluThrThr 343
Db 262319 GGTTCFGG-----AATCTTTAAAGTCGATATTATTAATAAATGATTATCAAAAATAC 262372
QY 344 LeuAlaIleGlyIleSerGlnProArgAsnTyrArgGlyAsn----- 357
Db 262373 CTGTGATATCAACAAGTATTCATATTTTATTAATAATOCAGATTAAATGCCAGA 262432
QY 358 TyrThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAla 377
Db 262433 TTTTTTACAATGATTTTAAATATAATTTTGACAATATTTTCGAATATTATATAAAATACA 262492
QY 378 PheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 262493 TAT-----GGTTTGAAGGTAATTTGGGA 262516

RESULT 7
US-09-886-468-6
; Sequence 6, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2950
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(2866)
US-09-886-468-6

Alignment Scores:
Pred. No.: 0.00341 Length: 2950
Score: 116.00 Matches: 131
Percent Similarity: 30.23% Conservative: 83

Best Local Similarity: 18.50% Mismatches: 226
Query Match: 3.68% Indels: 288
DB: 1.0 Gaps: 29
US-09-857-669-2 (1-609) x US-09-886-468-6 (1-2950)
QY 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
Db 380 TATTTTAAATAATTTCTCAGGAACCTACAAGGAAGGGGCT-----GTACTTTGTTGC 433
QY 128 ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSer 147
Db 434 CAAGATCCTCAAGCAACGCGCGTCTTTCTGGGTCTCCACGCTCTCTTTTATTTCAGAGC 493
QY 148 AspGlyAsnLeuAlaGlu-----TyrTyrArgAsnAlaLeu----- 159
Db 494 CCGGAGATATTAAAGACAGGATGCTCTATTCAAAAATGCACTTATGCTCTTAAC 553
QY 160 -----GluAsnTrpGlnGlnProValGly----- 167
Db 554 AATTATGTAGTCGTTTTTGAACAAACCAAGTAAGACTAAAGCGCGAGCTATTATGTGG 613
QY 168 -----SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSer 181
Db 614 GCGAATGTTTACTATAGTACGCACTACGATTCCGCTCTCTTTCTATCAGAATCAGCCACT 673
QY 182 ValLeuGlyAlaValThrArgLysGlyTyrProLeu----- 193
Db 674 TTTGGAGGTGCTATCCATCTCTTCAGGT---CCCCACAGATTCAGTAATAATCAGGCAGAG 730
QY 194 -----AlaLysLeuGlyAsnThrArgAlaAlaValAsnProAsp 206
Db 731 ATAAGATTTGCACAAANTACTGCCAAGAAATGCTCTCGAGGGGCTTTGTACTCC----- 784
QY 207 ThrAlaThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAsp 226
Db 785 -----GATGGTGATATTGATATTGATCAGAAATGCTTATGTTCTATTTCGAGAA 832
QY 227 PheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPhe 246
Db 833 AATGAGGCATTGACTACTGCTATAGTAAGGAGGGGCTGCTGTTGTTCTTCCACTTCA 892
QY 247 GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAlaLeuGluGln 266
Db 893 GGAAGTAGTACTCCAGTTCTTATTTGACTTTCTCAGCAATAAACAGATTAGTCTTTTGA 952
QY 267 AsnGlyHis-----TyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
Db 953 AGAAACCATTTCCATAATGGGTGCGGAGGCCATTTATCTAGG----- 994
QY 285 AspArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGly 304
Db 995 -----AAACTTAGCATCTCTTTCAGGAGGTCTACTCTATTTTATCAATAAT 1039
QY 305 IleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAsp----- 320
Db 1040 ATATCATATGCAAAATTCGCAAAATTTAGTGGAGCTATTGCCATTGATATCTGAGGGAG 1099
QY 320 ----- 320
Db 1100 ATCAGTTTATCAGCAGAGAAAGGAACAATTACATTTCCAAGAACCGGAGGCTTACCG 1159
QY 320 ----- 320
Db 1160 TTTTGAATGGCATCCATCTTTTACAAAATGCTAAATTCCTCTGAAATTAACAGCGAGAAAT 1219
QY 321 -----TyrTyrAsnLeuPhe----- 325
Db 1220 GGATACTCTATAGAATTTTATGATCCCTATTACTTCTGAACAGATGGGTCTACCCAAATG 1279
QY 326 -----AsnLysGlyTyrIleGlySerValValTrpAspMetAsp 338
Db 1280 AATATCAACGGAGATCCTTAAAAAATAAGAGTACACAGGGAGCATACTCTTTCTCGAGAA 1339

QY 191 TyrProLeuAla-----LysLeuGlyAsnThrArgAlaAlaValAsnPro 205
|||||
Db 2104 TATCCGGTGTCTGTCGCCCATCTATCTCGCGGAGCGTGCATGAGATCGCTCTAC 2163
QY 206 AspThrAlaThr-----ValAspLeuAsnValValValAspSerGlyArgPro 221
|||||
Db 2164 CATGAAGCAGAGCTTCTCCATTATGGAATTCAGTAAATAATCAATCCGCGACCCCT 2223
QY 222 IleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSer 241
-|||
Db 2224 GTCCGTATC---GATAAATATTTGACCGAAAGAAATAGAGTGTGATCCCTTCTGAC 2280
QY 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
|||||
Db 2281 GGTGAACGGTCTCATTCGGGG-----ATTATG 2310
QY 262 GlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
-|||
Db 2311 GAGCATCGAACGCGCGGGCTCCATCGGGGATTCGATT----- 2352
QY 282 LeuGlnGlyAspArgValProValValValSerValThrGluValLysArgHisLysLeu 301
-|||
Db 2353 -----GCCGTTTACCCGCCAATCTCTTCAGAGGACATCAAGAAAGCTT 2400
QY 302 GluThr---GlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIleAlaTyrAsp 320
-|||
Db 2401 GAAGCATACAGATTAGCTAGCAAGGGCTTAAACATCGTCGCGCTCAACATTCAG 2460
QY 321 TyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMetAspLysTyr 340
-|||
Db 2461 TTTGTCTTTCAAAAGATGAGTTTACGTCTCTGAAGTC----- 2499
QY 341 GluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThr 360
-|||
Db 2500 -----AACCCGAGATCGAGCGCAACGGTCCGCTGCTTCTC 2532
QY 361 SerAsnValSer-----TyrAsnArgSerThr----- 369
Db 2533 AGCAAAATCACCGGCATTCGATGCGCAATCTCGCAAAAGATGATTCTCGGAGGGAAG 2592
QY 370 -----ThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrp 383
-|||
Db 2593 CTTGCCGATATGGGTATACAGAGAGCGCTCCAAAAGAG-----CAGGAAGCGCTA--- 2643
QY 384 TyrValArg-----AspArgAlaGlyIleAspAlaArgLeuGly 396
-|||
Db 2644 TATGTCAAAGTCCCGGTCTTCTGTTGCGAAGCTTAGAAGAGTGGATATCAGCTCGGA 2703
QY 397 AlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp----- 410
-|||
Db 2704 CCGGAATGAAATCGACCGGG---GAAGTCATGGCGAAACAGACACATTGGAAAAAGCG 2760
QY 411 -----IleAspLeuGlyAsnSerHisAlaThrMetLeu 421
-|||
Db 2761 CTATATAAGGACTGATCGCTTCAGGCATTCAMATTCGGAACCTACGGCTCTGTCTCTCG 2820
QY 422 ThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyr 441
-|||
Db 2821 ACAGTTGCTGATAAAGATAAA---GAGGAAGGGCTGTGTATCGCCAGCGGTTCATCGG 2877
QY 442 LeuAspGlyLysIleGlyThrThrLeuGlyThr-----PheLeuSerSerThrAlaLeu 459
-|||
Db 2878 ATCGGTATATAAATCTTGCTAGCGAAGGAGCGCAGGCTATTAAAGATGCATCCGTA 2937
QY 460 ---IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGly 478
-|||
Db 2938 CTGTTTAAACCGTCGGGAAATATCGGC-----GAAGAAGGTACAAATCTCTT 2985
QY 479 ThrPheIleAlaGlyGlnAlaGlyTyrThrVal----- 490
-|||
Db 2986 GATGTCTATTCGAACGCGGAGCGAGTTCTGCTCATACGCTGACAAAGGAAGACAG 3045

QY 491 ---AlaArgAsp 493
|||||
Db 3046 CCGCAAGGGAC 3057
RESULT 10
US-09-332-226-1
: Sequence 1, Application US/09332226
: Patent No. US2002025318A1
: GENERAL INFORMATION:
: APPLICANT: Sparling, P. Frederick
: APPLICANT: Cornelissen, Cynthia N.
: TITLE OF INVENTION: Transferrin-Binding Proteins From
: TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ImClone Systems Incorporated
: STREET: 180 Varick Street
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10014
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/332,226
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/363,124
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/973,336
: FILING DATE: 05-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/572,187
: FILING DATE: 23-AUG-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Gallagher, Thomas C.
: REGISTRATION NUMBER: 37,066
: REFERENCE/DOCKET NUMBER: SPA-1-PDC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 645-1405
: TELEFAX: (212) 645-2054
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3286 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: Neisseria gonorrhea
: STRAIN: FA19
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 406..3150
US-09-332-226-1
Alignment Scores:
Pred. No.: 0.0115 Length: 3286
Score: 112.00 Matches: 150
Percent Similarity: 34.09% Conservative: 89
Best Local Similarity: 21.40% Mismatches: 281
Query Match: 3.53% Indels: 181
DB: 10 Gaps: 36
US-09-857-669-2 (1-609) x US-09-332-226-1 (1-3286)

Db 29603 ATCATTGTTAAAGCTGGGCGGAATCTTCGTTTGAACAGGCGGTAACCGCC----- 29656
Qy 513 GlyTyrGluLeuAspSerIleGluLeuAlaGlyProAsnGlySerValLeuPro----- 530
Db 29657 -----GGCACCGTTAACCTTCTGCGCGCGGAGGACACCAACT 29695
Qy 531 -----GluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArg 547
Db 29696 CTCACCTGCAGAACCCCTCGAGCCTCCCAACTCTTCACTGTCCCTATCG--- 29752
Qy 548 ThrLeuSerGlyAlaValPheHisAsp-----MetGlyAspAla--- 560
Db 29753 -----GAGGCTGTGGTGTGCATACCAACGCCCTTCTCTCCAACTGGGAGACGCATG 29806
Qy 561 AlaAlaAsnPhelYsArgMetLysLeuLysHisGlySerGlyLeuGlyValArg 578
Db 29807 GAAGTAATCAGCAGCGACTTACTTTAGACTAGCTCGGTTTGCAAAATGCCT 29860

RESULT 12

US-09-864-761-4449/c
; Sequence 4449, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ IDS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4449
; LENGTH: 1969

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005366.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
US-09-864-761-4449
Alignment Scores:
Pred. No.: 0.013 Length: 1969
Score: 108.50 Matches: 135
Percent Similarity: 30.84% Conservative: 75
Best Local Similarity: 19.82% Mismatches: 184
Query Match: 3.42% Indels: 287
DB: 10 Gaps: 35

US-09-857-669-2 (1-609) x US-09-864-761-4449 (1-1969)

Qy 25 AspleuSerGluAsnLysAlaAlaGly-----PheAlaLeuPheLysAsnLysSerPro 42
Db 1868 GAGATTAGCGAGCCCTGGCTCCGGGAGCGCTTTCGCTCGAGAGCGCGCAGCATCC 1809
Qy 43 AspThrGluSerValLysLysLys-----ProLysPheProVal 55
Db 1808 GATGTGGGAAGCAACTCTTTACAAACCTATGAGCTGAGCGCGAATGAATCTTTGCGCTT 1749
Qy 56 ArgIleAspThrGluAspSerGluIleLysAspMetValGluGluHisLeuProLeuIle 75
Db 1748 CGCTGCACACGCGGAGGACAGCACCAG-----TACGCGGAGCTGGTG 1704
Qy 76 ThrGlnGlnGlnGluValLeuAspLysGluGln----- 87
Db 1703 TTG-----GAGCGCGCCCTGGACCGACAGCGGAGCTAGCTCTCCAGTTAGTGCTG 1653
Qy 88 -----ThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThr 101
Db 1652 ACGCGCTTGGACGGAGGACCCAGCTCTCTCCGCGACCTGCTTATTCACATCAGGTG 1593
Qy 102 Met-----LeuArgSerLysGlyTyrPheSerSerLysValSer 114
Db 1592 CTGGACGCGAATGACAATGCGGCTGTCTTCAACAGTCTTGTACCGCGCGCTCCTG 1533
Qy 115 LeuThrGluLysAspGlyAlaTyrThrValHisIle-----ThrProGly 129
Db 1532 GAGGATGCACCTCCGCGACGCGCGTGTGTACAGTCTTTCACACGGATCTGGATGAGGC 1473
Qy 130 ProArgThrLysIle-----AlaAsnValGlyValAla 140
Db 1472 CCCAAGCGTGAATATTACTCTTCGGCAGCCACCAACCGCGCGCTGCGCGCACTA 1413
Qy 141 IleLeuGlyAspIleLeuSer-----Asp 148
Db 1412 TTGCGCTTAGACCTTGTAAACGGGATGCTGACATCAAGGTCGGCTCGGATTCGAGGAC 1353
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 1352 ACCAACTCCATGAGATTTATCATCCAGGCCAAAGAGCGCGCAATCCCGAAGGACA 1293
Qy 169 -----AspPheAspGlnAspSerTrpGluAsnSerLys 179
Db 1292 CATTCGAAAGTGTGGTGGAGTTGTGGATGTGAATCACAACGCCCGCGGAGATCACAGTC 1233
Qy 180 ThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199
Db 1232 ACCTCGGTGTACAGCCCACTAGCCCGAGGAT-----GCCCTCTCTGGGAGTGTGTC 1185

QY	200	ArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly	219
DB	1184	ATCGCTTTGGTCTC-----AGTGAGTACGACCTG-----GAUGCTGGG	1149
QY	220	ArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIle	239
DB	1148	GAG-----1146	
QY	240	ValSerGlyLeuAlaAargPheGln-----ProGlyThrProTyrAspLeuAspLeuLeu	257
DB	1145	---AACGGCTGGTGACCTGGCGAAGTTCCACGGGTCTCCCTTTACGCTTACTTCTTC	1089
QY	258	LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAla	277
DB	1088	CTCAAG-----AATTACTTACTTTGAAAAACCACTGCA	1056
QY	278	AspPheAspArgLeuGlnGlyAspArgValPro-----	288
DB	1055	GACCTGGATCGG-----GAGACTGTGGCAAGATAACCTCAGCATCACCGCCGA	1005
QY	289	-----VallysValSerValThrGluVal	296
DB	1004	GACGCCGAACCCCTCCCTCTCAGCCCTTACAATAGTGGTGTTCAGTGTCCGACATC	945
QY	297	LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyClyLys	316
DB	944	AATGACAAACCTCCACAACT-----TCTCAA	918
QY	317	IleAlaTyrAspTyrTyr-----AsnLeuPheAsnIysGlyTyrIleGlySer	332
DB	917	TCTTCTACGACGTTTACATTTGAAGAAACAACCTCCCGGGCTCCAATACTMAACCTA	858
QY	333	ValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArg	352
DB	857	AGTGTCTGGGACCCGAC-----GCCCGCGG	831
QY	353	AsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsn	372
DB	830	AATGTCGGCTTTCTTTCTTCTC-----	807
QY	373	LeuGluLysArgAlaPheSerGly-----GlyIleTyrTyrValArgAspArg---Ala	389
DB	806	TTGGAGCAGGAGCTGAACCGGCTAGTGGTGCCTATTTCACAATAAATCGTCACAA	747
QY	390	GlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySer	409
DB	746	GGCATAGTGTATCTCTTA-----GTGCCC-----	723
QY	410	AspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu	429
DB	722	--CTAGACTATGAGGATCGCGGGAAATTTCAATTAACACT-----	684
QY	430	LeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThr	449
DB	683	-----CATATCAGCATGGG-----	669
QY	450	LeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhe	469
DB	668	-----GGCACCCTGGTCTCTCCACCAACATCAGCTGAGCAATATTT	627
QY	470	PheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThr	489
DB	626	GTCACCTGATCGCAAT-----	612
QY	490	ValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAla---	508
DB	611	-----GACAATGCC-----CCCCAGCTCTATATCTCGGCCAGGTGGGAGCTCG	567
QY	509	-----SerSerValArgGlyTyr	514
DB	566	GTGGAGATGTCTCGAGGTACTCTACGTGGCCACCTAGTGTACGGGTGGTAGCTGG	507
QY	515	GluLeuAsp-----SerIleGlyLeuAlaGly---ProAsnGly	526

```

Db      506  GACGGCGATGCAGGCGACAAATGCGTGGCTCTCTCCTACAGTCTCTTGGGATCCCTTAACCGAG  547
Qy      527  SerValLeu-----||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      446  AGCGCTTTTGGCCATAGGCGTGCACTGGTGCAAAATCAAGTACTGCCCGTCCAGTCCAAGAC  387
Qy      530  -----ProGluArgAlaLeuVal-----GlySerLeuGluTyrGlnLeu  543
Db      386  ACAGATTACCCAGCGCACTCTCACGGCTTGTGATCAAGACAAATGGGGAGCCTTCGCTC  327
Qy      544  ProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAsn  563
Db      326  TCCACCACTGCTTACCCTCACTGTGTCTAGTAACCGAGGACTCTCTCTGAAGCCCGAGCCGAG  267
Qy      564  Phe 564
Db      266  TTC 264

RESULT 13
US-09-864-761-21197/c
; Patent No. 21197, Application US/09864761
; Sequence US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21197
; LENGTH: 2004

```

```

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21197
; LENGTH: 2004

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005366.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EST_HUMAN HIT: BF206078.1, EVALUAE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P33450, EVALUAE 4.00e-61
; OTHER INFORMATION: NT HIT: g111416576, EVALUAE 0.00e+00
US-09-864-761-21197

Alignment Scores:
Pred. No.: 0.0134 Length: 2004
Score: 108.50 Matches: 135
Percent Similarity: 30.84% Conservative: 75
Best Local Similarity: 19.82% Mismatches: 184
Query Match: 3.42% Indels: 287
DB: 10 Gaps: 35

US-09-857-669-2 (1-609) x US-09-864-761-21197 (1-2004)
Qy 25 AspleuSerGluAsnLysAlaLaGly-----PheAlaLeuPheLysAsnLysSerPro 42
      : : : : : | | | | | : : : : : | | | | |
Db 1764 GAGATTAGCAGGCGCGTGGCTCGGGGAGCGCGCTTTCGGCTCGAGAGCGCGCACGATCCC 1705

Qy 43 AspThrGluSerValLysLeuLys-----ProLysPheProVal 55
      | | | | | | | | | | : : : : :
Db 1704 GATGTGGGAGCAACTCTTTACAACTATGAGCTGAGCGCCGAATGAATACTTTGGCGTT 1645

Qy 56 ArgIleAspThrGlnAspSerGluLeuLysAspMetValGluGluHisLeuProLeuLe 75
      : : : : : | | | | | : : : : :
Db 1644 CGGCTGCAGACGGGGAGACAGACACCAAG-----TAGCGGAGCTGGTG 1600

Qy 76 ThrGlnGlnGlnGluValLeuAspLysGluGln----- 87
      | | | | | | | | | | : : : : :
Db 1599 TTG-----CAGCGCGCCCTGGACCGAGACGGGAGCCCTAGCTCCAGCTTAGTCTG 1549

Qy 88 -----ThrGlyPheLeuAlaGluAlaProAspAsnValLysThr 101
      : : : : : | | | | | : : : : :
Db 1548 ACGCGCTTGGACGGAGGACCGCAGCTCTCCGCGACGCCCTGCCCTATTTCACATCAAGGTG 1489

Qy 102 Met-----LeuArgSerLysGlyTyrPheSerSerLysValSer 114
      : : : : : | | | | | : : : : :
Db 1488 CTGACGCGGAATGACAAATCGGCTCTCTCAACCAAGTCTTGTACCGGCGCGGCTGCTG 1429

Qy 115 LeuThrGluLysAspGlyAlaTyrThrValHisIle-----ThrProGly 139
      : : : : : | | | | | : : : : :
Db 1428 GAGGATGCACCCCGCGCAGCGCGCTGGTACAAAGTCTTGCACCGGATCTGGATGAAG 1369

Qy 130 ProArgThrLysIle-----AlaAsnValGlyValAla 140
      : : : : : | | | | | : : : : :
Db 1368 CCCAAGGTGAATATTATTACTCTTCGCGCACCCACACCGCGCGCTCGCGCACTA 1309

Qy 141 IleLeuGlyAspIleLeuSer-----Asp 148
      : : : : : | | | | | : : : : :
Db 1308 TTCGCTTAGACCTTGTAAACCGGGATGCTGACAANTCAAGGTTCGGCTGGACTTCGAGGAC 1249

Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
      : : : : : | | | | | : : : : :
Db 1248 ACCAACTCCATCAGATTACATCCAGGCCCAAAGACAGGGCGCCAAATCCCGAAGGACCA 1189

Qy 169 -----AspPheAspGlnAspSerTrpGluAsnSerLys 179
      : : : : : | | | | | : : : : :
Db 1188 CATTGCAAGTGTGTGTGGAGTGTGGGATGTGAATGCAACGCCCGCGGACATCAGCTG 1129

Qy 180 ThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199

```



```
QY 354 ---TyrArgGlyAspTyrThrSerAsn----- 362
Db 3448 TTAGTTAACCCAAATCAGGAACAGCAGTTCTGTACACGAGATTGGTATTTGATGGTAAA 3507
QY 363 -----ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPhe----- 378
Db 3508 GGTATGTTTATTAT-----TCAACGAGTGGTAAACAGCTAAAAAGCTTTTCATTAGC 3561
QY 379 SerGlyGlyIleTyrThrValArgAspArg-----AlaGlyIleAspAla 393
Db 3562 TTAGGAANAATGTTGATTATTTCGCATAATAACGGTTATATAGCTCACTGCTCAATCA 3621
QY 394 ArgLeuGlyAlaGlu-----PheLeuAlaGluGlyArgLysIleProGlySerAspIle 411
Db 3622 ATTAACGGTGCCTAATTAATTCTTCAATGGTATTCATTAAGAAATGCTATTAT 3681
QY 412 AspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsn 431
Db 3682 GATAATGGTAATAAAGTATGCTTATTATGGAATGATGGCGTGGTTAT----- 3732
QY 432 AsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGly 451
Db 3733 -----GAAATGGTGTACTATCTCTTGGTCAACAATGGCGTTATTTCCAA 3777
QY 452 ThrPheLeuSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThr 471
Db 3778 ANTGGTATTATGGCTGCGCTTAACAGTGTTCATCGTCTTCATATTTTGTATGCT 3837
QY 472 ProGluAsnLysLysLeuGlyThrPheIleIleArg-----GlyGlnAlaGlyTyrThr 489
Db 3838 TCTGGTGTCCAAAGCTAAAGGACAGTTATTATCAACTGCTGATGGAAGCTGCTTAC--- 3894
QY 490 ValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyAlaSer 509
Db 3895 TTGTATAGAGACTCAGGNAATCAAAATTTCAATCGTTTGTGTAGAANT-----TCC 3945
QY 510 SerValArgGlyTyrGluLeuAspSerIleGlyLeuAla----- 522
Db 3946 AAGGAGAATGGTCTTTATTATGATCACAATGGTGTGCTGTAAACCGGTACTGTAACGTT 4005
QY 523 -----GlyProAsnGlySerValLeuProGluArgAlaLeuLeu 535
Db 4006 AATGGCAACAGCTTCTACTTTAAACCTTAATGGT-----GTTCACGCCAAGCAGAAATTT 4059
QY 536 Val-----GlySerLeuGluTyrGlnLeuPro----- 544
Db 4060 ATCAGAGATGCAAAATGGATATCTAAGATATTATCATCTAATTCGGAAATGAAGTTGCT 4119
QY 545 -----PheThrArgThrLeuSerGlyAlaValPhe---HisAspMetGlyAspAlaAla 561
Db 4120 ANTCCCTTTGTAGAAATCCAAAGGAGAATGGTTCTTATTGATCACAATGGTATCCCT 4179
QY 562 AlaAsnPhelLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPhe 580
Db 4180 GTAACGTGGCCAGAGTTGTTAATGGACACGCTCTATTTTAAGTCTAATGGTGTTC 4236

RESULT 15
US-09-815-242-8513
; Sequence 8513, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
```

```
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8513
; LENGTH: 9477
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(9477)
US-09-815-242-8513

Alignment Scores:
Pred. No.: 0.147 Length: 9477
Score: 108.50 Matches: 91
Percent Similarity: 34.81% Conservative: 66
Best Local Similarity: 20.18% Mismatches: 159
Query Match: 3.42% Indels: 135
DB: 10 Gaps: 20

US-09-857-669-2 (1-609) x US-09-815-242-8513 (1-9477)
QY 27 SerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGluSer 46
Db 3010 TCTTTAAATACAGCAATGGTAACTTAATT-----AATCGGATTGCAGATCATCAAGCC 3063
QY 47 ValLysLeuLysProLysPheProValArgIleAspThr---GlnAspSerGluLys 65
Db 3064 GTTGACCAACCGTGGTAACTTC---ATCAATGCTGATACTGATAACAACTGCTTATAT 3120
QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGln----- 79
Db 3121 ACAGCGGTAATGAAGCAGAGCCATGATTAAACAAACAACTGCTCAAAATGCGAACCAA 3180
QY 80 -----GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaPro 96
Db 3181 CCAGAAAGTAGAACAGCTATTACTAAGTTCAAACTACACTTCAAGCGTTAANTGGAGAT 3240
QY 97 AspAsnValLys----- 103
Db 3241 CATAACTTACAAGTTGCTTAAACAAATGCGAGCGCAAGCAATGCTGTTTAAACAGCTTA 3300
QY 104 ArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThr 123
Db 3301 ANTGATCTCTCAAAAACAGCATTTAAAGACCAAGTTACTGCTCCAACTTTAGTAACGCA 3360
QY 124 ValHis---IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeu 142
Db 3361 GTTCATCAATTTGAACAAAATGCGAATACG-----CTTAAACCAACCAATGCTGCTTA 3414
QY 143 GlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrp 162
Db 3415 AGAGAAAGTATTCAGAT---AACGCGAGCAACTAAACCAATAGCAATATATATCAACGNA 3471
QY 163 GlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerVal 182
Db 3472 GATCAACCCAGAACAAACAACTAGCATCAA-----GCAGTG 3507
```

```
QY 183 LeuGlyAlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAla 202
Db 3508 CAAGCTGCA-----AATAGCATT 3525
QY 203 ValAsnProAspThrAlaThrValAspLeuAsnValVal-----Val 216
Db 3526 ATCAATGAGCAACAGGACATAGATAAATGCGATTAAATCAAGACGACGAACTGTG 3585
QY 217 AspSerGlyArgProIleAlaPheGlyAspPheGluIleThrClyThrClnArgTyrPro 236
Db 3586 AATACAAAGAAAGCAGCATTCATGGTGATGTGAAGTTACAAATGATTAAGATCATGCT 3645
QY 237 GluGlnIleValSerGlyLeuAlaPheGlnProGlyThrProTyrAspLeuAspLeu 256
Db 3646 AAACAACGGGTAGTCAATTAGCATCTAAACAATGCACAAAACATATGAAGATACG 3705
QY 257 LeuLeuAspPheGlnGlnAlaLeuGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
Db 3706 TTGATTGATGAACAACT----- 3726
QY 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
Db 3727 -----AGAATAGCAGTTAAGCAAGATTGTGACTCAAGCA 3759
QY 297 Lys-----ArgHisLysLeuGluThrGlyLe----- 305
Db 3760 CARGCATTAGATCAACTATTGATACATTACAAACAAGTATTGCTGACAAAGATGCAACA 3819
QY 306 ArgLeuAspSerGluTyr-----GlyLeuGlyGlyLysIleAlaTyrAspTyrTyr 322
Db 3820 CGTGAGCAGTGCGATATGTCAATGCAGAACCGCAATAAATAAACAAGCATATGAT----- 3873
QY 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValThrAspMetAspLysTyrGluThr 342
Db 3874 -----GAACGAGTTCAAAATGCTGAG 3894
QY 343 ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
Db 3895 TCTATCATTTGAGGATTAATAATCCGACTATCAATAAAGGAAAT-----GTATCAAGT 3948
QY 363 ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIle 382
Db 3949 GCGACACAGCAGTGACACATCATCAAAAT----- 3978
QY 383 TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu 402
Db 3979 -----GGTTTAGAT-----GGTGTTCAGGATTAGCTCAA 4008
QY 403 GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr 422
Db 4009 GATAGCAACACAGCTGGAATTCCTG-----ATCATTTAGATCAATTACA 4056
QY 423 AlaSerTrpLysArgGlnLeuLeuAsnAsnVal 433
Db 4057 CCAGCTCAACAACAAGCGCTTAGAAAATCAAATT 4089
```

RESULT 16

```
US-10-044-090-98
; Sequence 98, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044, 090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 4065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 118619.8
; NAME/KEY: unsure
; LOCATION: 3146
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-98
```

Alignment Scores:

```
Pred. No.: 0.0454 Length: 4065
Score: 108.00 Matches: 130
Percent Similarity: 32.08% Conservative: 83
Best Local Similarity: 19.58% Mismatches: 271
Query Match: 3.40% Indels: 180
DB: 12 Gaps: 29
```

US-09-857-669-2 (1-609) x US-10-044-090-98 (1-4065)

```
QY 41 SerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThrGln 60
Db 2007 GCACCTGACGAGAAATGAAGACCACCATCCAGAGAGGCTTGT-----GCTCAG 2057
QY 61 AspSerGluIleLysAspMetValGlu-----HisLeuProLeuIleThrGlnGln 78
Db 2058 AATGAGAGCTGCAGTGAAGAACTACACCGACTTCATTACCACTGTATTCAAGAAGA 2117
QY 79 -----GlnGluGlu-ValLeuAspLysGluGlnThrGlyPheLe 91
Db 2118 GGGCAAGCGCTGTTGACTGCAGGAAGACGCTGCTGGTGCATGCAGCAGGCTGGGC 2177
QY 91 uAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSe 111
Db 2178 ACCCTCTCCATTGTGATAAAC-----TTTGGAAAC 2207
QY 111 rLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrProGlyPro-A 131
Db 2208 CAAAATCTCCACGAGAGAGGCTGGGAGGAGCAGATGTGTGTCAAACTCTCGGAGAAC 2267
QY 131 rgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnL 151
Db 2268 GTCCCGGAAAAAGGCTGAATATTATTATGTGGCTGAA----- 2308
QY 151 euAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheA 171
Db 2309 -----GGAGCAATGTATACCCAAATAAACCCATCACCTCT----- 2344
QY 171 spGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyT 191
Db 2345 -----CAGAAATCAAGAGCTTGTG-----GTCACGACGCTGGGCT 2381
QY 191 yProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValA 211
Db 2382 AT-----GACACACGCTGTGAACA 2399
QY 211 spLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrG 231
Db 2400 TCCTCGGCACGTCGACGAGAGAGGACCCCTCGCGCATTTGAC-----AGGATCTTGG 2453
QY 231 lYThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrP 251
Db 2454 CCAGCCGCATGGGAGTGGAGCAGTCATCGCC---TTGCTAGAGGTACCCCGGACAC 2510
QY 251 ro----- 251
Db 2511 CAGCTTGCCTGCTGTCTACTGAACGGGAACCGCGTGGCGCTGCTGCTGATGAGTGGCG 2570
QY 252 -----TyrAspLeuAspLeuLeuLeuAsp-----PheGlnGlnAlaLeuG 265
Db 2571 TGCAGATGACTCAGGATGTGCAAGAGCGGATGGACAGAGGAGATTTCAGGATGCGGTT 2630
QY 265 luGlnAsnGlyHis-----TyrSerGlyAlaSerValGlnA 277
Db 2631 GACTCCGAGGAGGAGGCTTTGCGGGCAACCTGAACACCTACAGCGACTTGCATC----- 2686
```

```
QY 277 leAspPheAspArgLeuGlnGlyAspArgValPro-----ValLysValSerValT 294
Db 2687 -----NAGTGGCGGATGATCAGATCCCAAGACCAATTCGAACGTAGCTGTCA 2735
QY 294 hrGluVal-----LysArgHisLysLeuGluThrGlyIleArgLeuA 308
Db 2736 TCAACGTGGGGCACCGCGGTGGATGAACGACGCGTACGCTGCGGTGG 2795
QY 308 spSerGluThrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLysPheAsnLysG 328
Db 2796 GCATTCGCCAGCCACAGGATGCTGCCATCTATGATGGCTTTGACGGCTTTGCCAAGG 2855
QY 328 lyTyrIleGlySerValValTrp----AspMetAspLysTyrGluThrThrLeuAlaIaG 347
Db 2856 GCCAGATCAAGAAATCGCTGCGACAGATGTCGGGGCTGCGACCGCGGAGGTCCA 2915
QY 347 lyTyrSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnA 367
Db 2916 TTTCTGGGCAAAACCGCTTCTCCCGGGAAGTACTTGGAGAGATCGCCACAGATGC 2975
QY 367 rgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTyrValArgA 387
Db 2976 GC-----ACGCACAGCATCAACGCGCTGCTGATCATCGGT----- 3010
QY 387 spArgAlaGlyIleAspAlaArgLeuGluLysArgAlaPheLeuAlaGluGlyArgLys---- 405
Db 3011 -----GGATTGAGGCGCTACCTGGGACTCTGGAGCTGTCAGCGCGCGGAGAACG 3062
QY 406 -----ACGAGAGTCTGTCTCCCGCATGTGTCATGTTCCTCGTACTGTCTCCAAACATGTCCGG 3122
Db 408 lySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTyrLysArgG 428
Db 3123 GTTCGATTTTACCATCGCGGCANACCGCCCTGAACATCATACCGACACCTGGGACC 3182
QY 428 lnLeuLeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyT 448
Db 3183 GCATCAAGCAGTCCCGCAGCGGAACCAACGCGCGGTGTTTCATCATCGAGACCATGGCG 3242
QY 448 hrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyT 468
Db 3243 GCTACTGTGCG---TACCTGGCCACATCGGGGGCTGCGCGCTGGAGTGTGCGGCAT 3299
QY 468 yPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArg-----GlyG 485
Db 3300 ACATTTTCGAAGAG-----CCCTTCGACATCAGGATCTGCAGTCCA 3341
QY 485 lnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgS 505
Db 3342 ACCTGGAGCACCTGACGGAGAAATGAAGACCACTCCAGAGAGCGCTTGTGCTCAGAA 3401
QY 505 erGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle----GlyLeuAlaGlyP 524
Db 3402 ATGAGAGCTGCAT-----GAAAACTACACACCGCATTTTCATTTACCAAGCTATTTCAG 3455
QY 524 roAsnGly---SerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyr---- 541
Db 3456 AAGAGGCAAAAGCGCTTTTACCTGCAGGAAGACCTGCTGGCTCATCATCAGCAGCGGTG 3515
QY 542 -----GlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspA 560
Db 3516 GGGCACCTCTCCATTTGATAGAACTTTGGAAACCAAAATC-----TCTGCCA 3563
QY 560 laAlaAlaAsnPheLysArgMetLysLysHisGlySerGlyLeuGlyValArgTyrTp- 579
Db 3564 GAGCTATGAGTGGATCACTCAAAACTCAAGAGAGCGCGGCGCAGAGGAAAAATTTA 3623
QY 580 -----PheSerProL 583
Db 3624 CCACCGATGATTCATTTGCTGCTGGGAATAAGCAAAAGAAACGTTATTTTCAACCTG 3683
QY 583 euAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleArgTyrPheHisI 603
```

```
Db 3684 TGCAGAGCTGAAGAGCAACGCGATTTTGACACAGGATTTCCCAAGAACAGTGTGGTGC 3743
QY 603 leSerLeu 605
Db 3744 TCAAGCTA 3751
RESULT 17
US-09-954-456-1601
; Sequence 1601, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1601
; LENGTH: 14800
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-1601
Alignment Scores:
Pred. No.: 0,379 Length: 14800
Score: 107,50 Matches: 119
Percent Similarity: 35,17% Conservative: 85
Best Local Similarity: 20,52% Mismatches: 230
Query Match: 3,39% Indels: 147
DB: 10 Gaps: 29
US-09-857-669-2 (1-609) x US-09-954-456-1601 (1-14800)
QY 64 IleLysAspMetValGluGluHis---LeuProLeuIleThrGlnGlnGluGluVal 82
Db 12286 ATCTCCACCCCTGAGGAGGAGCCGCGCTGCCGTGCGAGGTGAGGCGGCGGCTACAGCGCGGCTC 12345
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
Db 12346 TTCGATGAGGAGATCAAGCAGATCTGTACCCGCCCTCGGACGAC----- 12390
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
Db 12391 -----ACCAAGGGCTTCTTTGACCCCTAAC-----ACGAGGAGAACCTCACCTAC 12435
QY 123 -----ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly--- 138
Db 12436 CTCACGCTGATGAGCGTGTGTATCACTACCCCGACGCGGCGCTGTCTCTTTCGCGCTG 12495
QY 138 ----- 138
```



```
Db 250 GACCAGTACACATTTCAAAAAGAGTACTCTGCACAACAGGCTC---CACCTTGTCTCATCG 306
QY 71 HisLeuProLeuIleThrGlnGlnGluGluValLeuAspLysGluGlnThrGlyPhe 90
Db 307 ATTTTCGACATCATTAAGAAGTGAAGAAAGAGCCAGCAAGAAAGAGAAAGAAAGCAAG 366
QY 91 LeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPheSer 110
Db 367 CAGGCCCATCTGAAAGACACCATCAATCTGTT-----CAAGAAAGCTGACT 417
QY 111 SerLysVal-----SerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 418 TCGATGTACGGACTCTGCTGTCCAGGATTCAGTCAACACCTGTGAGGCCAGCGAC 477
QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 478 GGAGAGCTTTCCATCAGCCAGGATTCGGTCATCACCCCGCTCAATCAGCTGATGAGCAA 537
QY 149 -----GlyAsnLeuAlaGluTyrTyrArgAsnAla----- 158
Db 538 GAATCCGGCGCAAGAAAGCTGATCAACCGCAAGAACATCAACGAAAGAACTGAAGGCG 597
QY 159 -----LeuGluAsnTrpGlnGlnProValGlySer----- 168
Db 598 AATTCGATACCTTCAAAATTAATAAACCGCGCTGATCAATCGGACGCTTTCGGATCAT 657
QY 169 -----AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 658 CCCAATTAGTATAGTATCCGTGCAAGACGAGCAGCAGGCGCAAGAGCGGCTGACAGC 717
QY 186 ValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 718 GTA-----CAGCAGGTACAGATCAAGCAG 741
QY 206 AspThrAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGly 225
Db 742 GGGCAGATACTCGTTGAGAAATCAAGCTGATCGACAGGAA-----GTCTATCGCA 792
QY 226 AspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg 245
Db 793 AAGCTCGAATTGACAGGG-----CTCTTGAGCGGAGCCCAACATC 831
QY 246 PheGlnPro-----GlyThrProTyrAspLeuAspLeuLeuAspPheGlnAla 263
Db 832 TTCAAACCTGTCGGCGGACTG-----CTGCTTTTAATCGCGCTTTTTCATCTCAGCT 882
QY 264 LeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGln 283
Db 883 TTGGTT-----TACTAT-----TTCGAAAGCAATAC 909
QY 284 GlyAspArgValProValLysValSerValThrCluValLysArgHisLysLeuGluThr 303
Db 910 AGCAGCGGATTCGAAAGAAACAGTCG-----CTCATGCTGTTTTCGTTTGATCACA 960
QY 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsn 323
Db 961 GGGATCGTTTAA-----ATCGTCATGCAAGTGCAGC 993
QY 324 LeuPheAsnLys-----GlyTyrIleGlySerValValTrpAspMetAspLysTyr 340
Db 994 CTGTTTCACGACTTGAATCGGCATATCGC-----TAT 1029
QY 341 GluThrThrLeuAlaIleGlyIleSerGlnProArg-----AsnTyrArgGlyAsn 357
Db 1030 CTCGTCCGATTCGGATGGCGCTGATGCTCAATTAACGTGATCAATGAAAGCGTGGCG 1089
QY 358 TyrTrpThrSer-----AsnValSerTyrAsnArgSerThrThr 370
Db 1090 ATCCCTGTCGAGCATGATCTCTCCATTTGGGAAGCATGATGTTTAACCGAGGCTGTGACA 1149
QY 371 GlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGly 390
```

```
Db 1150 GGCACGCTTCAAC-----TATGGAATCGAACGTATTTCTTAA----- 1185
QY 391 IleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAsp 410
Db 1186 ATCAGCTCGATCGCGGGTATATCTGTTTTTAGGGAAG----- 1221
QY 411 IleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeu 430
Db 1222 -----CATACGCGGCGGCAAAATTTCTCCAGCCCGGTTTTATTTCTCTCTGATC 1272
QY 431 AsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeu 450
Db 1273 AACATCGTTGTT----- 1284
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1285 -----CTGTTTACCCTCGAGCTGATTCAAAACACGCGCAACCCGGGCTT----- 1329
QY 471 ThrProGluAsnLysLeuGlyThrPheIleIleArgGlyGlnAla-----Gly 487
Db 1330 -----GAGATGGGTACATATCTTGTCTATGGGGCGGTGTCCGGCATCGGA 1374
QY 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMet-----PheArgSer 505
Db 1375 TCGTCAGTT-----CTTGTCTCGGACTCATGCGGTTTTTGAAGC 1416
QY 506 Gly---GlyAlaSerSerValArgGlyTyrCluLeuAspSerIleGlyLeuAlaGlyPro 524
Db 1417 GGATTTGGCATCTCTCCACG-----ATGAAGCTTCTCGAGGTTTCAAAATCGG 1464
QY 525 AsnGlySerValLeuProGluArgAlaLeuVal-----GlySerLeuGluTyr 541
Db 1465 AATCATCCGCTCCTC-----CGCAAAATCTGACGAAACACCGGCGCATCATCAC 1518
QY 542 GlnLeuProPheThrArgThrLeuSer----- 550
Db 1519 ACGGTT---ATGCTCGCAAGCTTATCTGAGCGGCATGCGAAGCGTTCGGAGCAACAGG 1575
QY 551 -----GlyAlaValPheHisAspMetGly 558
Db 1576 CTGCTGGCAAGGTCGGCGCCTATTACCATGATATATCGC 1614

RESULT 19
US-09-974-300-2620
; Sequence 2620, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2620
; LENGTH: 2140
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2620

Alignment Scores:
Pred. No.: 0.025 Length: 2140
Score: 106.50 Matches: 125
Percent Similarity: 36.02% Conservative: 103
Best Local Similarity: 19.75% Mismatches: 200
Query Match: 3.36% Indels: 205
DB: 10 Gaps: 38
```

[illegible]

COUNTRY: USA
 ZIP: 10014
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/332,226
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/363,124
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/973,336
 FILING DATE: 05-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/572,187
 FILING DATE: 23-AUG-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gallagher, Thomas C.
 REGISTRATION NUMBER: 37,066
 REFERENCE/DOCKET NUMBER: SPA-1-PDC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 645-1405
 TELEFAX: (212) 645-2054
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3537 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOPHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORGANISM SOURCE:
 ORGANISM: Neisseria meningitidis
 STRAIN: FAM18, FAM20, B16B6, group X and group W135
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 721..3450
 FEATURE:
 NAME/KEY: mat-peptide
 LOCATION: 793..3447
 US-09-332-226-3

Alignment Scores:
 Pred. No.: 0.0542 Length: 3537
 Score: 106.50 Matches: 151
 Percent Similarity: 35.03% Conservative: 90
 Best Local Similarity: 21.95% Mismatches: 285
 Query Match: 3.36% Indels: 162
 DB: 10 Gaps: 36

US-09-857-669-2 (1-609) x US-09-332-226-3 (1-3537)

Qy 85 ----LysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLe 103
 Db 957 GACCCGTTATGATCCGGGTATGTCGGTGGTCTGCAACAGGGTGGGGCGCAAGTTCCGGCTA 1016
 Qy 103 uArgSerLysGlyTyrPheSerLysValSerLysValThrGluLysAspGly----- 120
 Db 1017 TTCAATACGGCGCATGATAAAACCGCGTTTCTTACG---GTACACGGCGTTTCGCA 1073
 Qy 121 -----AlaTyrThrValHisIleThrProGlyThrProArgThrLysIleAlaSerValG 138
 Db 1074 AATACAGTCCTACACCGCGCAGCGGCATTCGGT---GGGACGAGGACGGCGGTAGCAG 1130
 Qy 138 yValAlaIle-----LeuGlyAspIleLeuSerAspG 149
 Db 1131 CGCGCGCAATCAATGAAATCGAGTATGAAACGTCAGGCGCTTGAATCAGCAAGGTTTC 1190
 Qy 149 yAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsn-----TrpGlnGlnProVa 166
 Db 1191 GAATTCATCAGAATACGGAACCGCGCATTCGCAGGTTCCGTCGTCATTTCAAAACCAAAAC 1250
 Qy 166 lGlySerAspPhe-----AspGlnAspSerTrpGlu---AsnSerLysThrSerValLe 183
 Db 1251 CGCAGCGGACATTCGGGAGAGGGAACAGTGGGGCATTCAGAGTAATAAATCCCTATTTC 1310
 Qy 183 uGly-----AlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrAr 200
 Db 1311 GGGAAAGACCATGCGCTGACGCAATCC-----CTTGGCGCTTCGCGGACGCGCG 1361
 Qy 200 gAlaAla-----ValAsnProAspThrAlaThrValAspLeuAsnValValVa 216
 Db 1362 CGCGCGGGAAGCCCTCTTATTATTACTAAACGGCGGGTGGGAAATCCATCCGCATAA 1421
 Qy 216 lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPr 236
 Db 1422 AGATCCGCGCAAG-----GGTGTGCAGAGCTTC-- 1449
 Qy 236 oGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLe 256
 Db 1450 -AACCGCGTGGTGGGACGAGCAAGAGGGTGGCAGTCAGTCAGATATTTTCATT 1508
 Qy 256 uLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlnPheIstYrSerGlyAlaSerValG 276
 Db 1509 G-----TGCAGAAAGAATGCCAATGGATATCGGCCCTGTGA 1547
 Qy 276 nAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVa 296
 Db 1548 A-----AACAAAGCTGAAAGAAGAT---GCCTCGGTCAAGATGAGCGCAAAACCGT 1595
 Qy 296 lLysArgHisLysLeuGluThrGlyIleArgLeu-----AspSerGluTyrGlyLe 313
 Db 1596 CACACGCGAGGATTTATCCGGCTCCACCGCTTACTTCCGACCCCGCTTGCATATTCGAG 1655
 Qy 313 uGlyGlyLysIleAlaTyrAspTyrAsnLeuPheAsnLeuGlyTyrIleGlySerVa 333
 Db 1656 CCAATCATGGCTGTTCCGACCGGGTTCGATTTGGACACCGCCATTTATTCGAGCGCT 1715
 Qy 333 lVal-----TrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyLeSe 349
 Db 1716 TCTCGAAGCTACGACGACAGACCTTTGATACACGGGATATGACTGCTCTCTCTATTATTAC 1775
 Qy 349 rGln-----ProArgAsnTyrArg-----GlyAsnTyrTrpThrSerAsn-- 362
 Db 1776 CAGTGAAGATTATGTACCGGTTTCGCTGAAAGGCTCTGGCAAAATATTCGGCGCATATAA 1835
 Qy 363 -----ValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAl 377
 Db 1836 GCGAGAAAGGCTGTTGTTTCAGGAGAGCGGACGATACATTCGAGGCTATC-----GG 1886
 Qy 377 aPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaLeuGlyAl 397
 Db 1887 TTACGGTACCGGGTGTGTTTAT-----GATGAACGCCCATACTAAAAACCGCTACGGGT 1940

```
QY 397 aGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHi 417
|||:|||||
Db 1941 CGNATGTTTACCAT-----AATCGTGAAGCATACCTGGCGCCGATTA 1985
QY 417 sAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeuHisProGl 437
|||:|||||
Db 1986 CGCCGCACTT-----TCCTATGACCGCAA-----2010
QY 437 uAsnGlyHisTyrLeuAspGlyLysIleGlyThrLeuGlyThrPheLeuSerSerTh 457
|||:|||||
Db 2011 ----GGTATAGATTGGACACCGTTTGCACGACAGCCATGCTCTCAGCAGCGTTCGA 2066
QY 457 rAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLe 477
|||:|||||
Db 2067 TAAAAATTCCCGTCCGACGCAATAAACCGTATCTCTTATATAATCCGACCGGATGAT 2126
QY 477 u-----GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaLeu 492
|||:|||||
Db 2127 TTATGAAGAAAGCCGAACCTGTTCT-----CAAGCAGTATTTAAAAAGCATT 2174
QY 492 gAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly-----507
|||:|||||
Db 2175 TGATACGGCCAAATCCGTCACAATTTGAGTATCAATCTAGGTCAGCAGCGCTTTAAGTC 2234
QY 508 -----AlaSerSerValArgGlyTyrGluLeu-----516
|||:|||||
Db 2235 GCNATGTCACACGCGATTATATCTTCAMAGCGCATTCAGCGCATATGTTGATAC 2294
QY 517 -----AspSerIleGl 520
|||:|||||
Db 2295 CCGGAAAAGCCTCGTTTCCCAACGGAAGCAAGACACCCGTATAGGCTGTCTATCG 2354
QY 520 yLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuValClySerLeuGl 540
|||:|||||
Db 2355 CAAGACCGGTCATACATACATCGCGATACCTGGT-----TTCCGCAATACAC 2402
QY 540 uTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetClyAspAl 560
|||:|||||
Db 2403 CTATACAGACTGCACACCGAGGATATCGCGCAACGGTTATTAT-----2448
QY 560 aAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPh 580
|||:|||||
Db 2449 -GCACCGTTCAAGACAATGTCGTTTGGGC-----AGTGGGC 2486
QY 580 eSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLysIleAr 600
|||:|||||
Db 2487 GGATGTCGGAGCAGCAGCATACGTTACGATTACCGCAGCAGCATTCGGAAGATAAG-----2541
QY 600 gTrpHisIleSerLeuGlyThr 607
|||:|||||
Db 2542 ----AGTGTCCTACCGCACT 2559
```

RESULT 21

```
US-09-912-020-139
; Sequence 139, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001DVI
; CURRENT APPLICATION NUMBER: US/09/912.020
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
```

```
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: E. Coli
US-09-912-020-139
```

Alignment Scores:

```
Pred. No.: 0.0162 Length: 1149
Score: 104.50 Matches: 79
Percent Similarity: 35.31% Conservative: 46
Best Local Similarity: 22.32% Mismatches: 142
Query Match: 3.29% Indels: 87
DB: Gaps: 18
```

US-09-857-669-2 (1-609) x US-09-912-020-139 (1-1149)

```
QY 232 ThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrPro 251
|||:|||||
Db 175 ACCTCAGCGAGCGACCAAAATATCATTTGGATTA-----AATGTGCCG 219
QY 252 Tyr---AspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyr 270
|||:|||||
Db 220 TTGCTGATATTACAGCTTCGCTGAATTACAGC-----TAT 255
QY 271 SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLys 290
|||:|||||
Db 256 TCCAATAATATATGCAAAAGCATCGGATCATTTACTCGCTTTTACGCTTAATGTTCOC 315
QY 291 ValSerValThrGluValLysArgHisLysLeuGluThrGlyIleArg-----306
|||:|||||
Db 316 TTCAGT-----CATTCGATGCGTACAGACAGCTAGTCGGCATTCGT 357
QY 307 ---LeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPhe 325
|||:|||||
Db 358 AATTCAACAGCGCATACAGTATGTCAACAGATTGGAAGCGGCATGACCAATCTA---414
QY 326 AsnLysGlyTyrIleGlySerValValTrpAspMetAsp---LysTyrGluThrThrLeu 344
|||:|||||
Db 415 ---TCGGGGTTATGGCACCTGCTGCGCGGATAATACCTGAATATATACGTTACAGTC 471
QY 345 -----AlaAlaGlyIleSer-----GlnProArgAsnTyr 354
|||:|||||
Db 472 GGTAAACCCAGCGAGGTAATACATCGTCTGGCACCATGCTTACAGTCTCTTAATATT 531
QY 355 ArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGlu 374
|||:|||||
Db 532 CGTGAGCTTATGTTAATAACTACTANTGCTCGGTTACAGTGGAGTGTGACACACCGAGATT 591
QY 375 LysArgAlaPheSerGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArg 394
|||:|||||
Db 592 TATTACGGAATGAGTGGTGGGATTATTGCTCATGCTGATGCAATCATCTTGGACACCGC 651
QY 395 LeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGly 414
|||:|||||
Db 652 CTGGCGCACACAAATGGTCTGTT-----AAGGCTCCTGGTGTGAT---ATGTCAAA 702
QY 415 AsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLeu 434
|||:|||||
Db 703 ATAGAGAACCAGACCGGAATTCATACCGAGCTGGGT-----GGCTATGCCATA 750
QY 435 HisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeu 454
|||:|||||
Db 751 TTACCAATTTGCCAGCAGATATAGAGAAAACCGTGTCTCTTACGCGGAATTCCTTGC 810
QY 455 SerSerThrAlaLeuIleArgThr-----SerAla 464
|||:|||||
Db 811 GATAAATGTTGAACCTGGATGAAACCGTGGTCTCATCTCCCACTACCGGTGATTTGCC 870
QY 465 ArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
|||:|||||
```

```

Db 871 AGACCAACATTT-----AATGCACAAATTCGGC-----GGG 900
Qy 485 GluAlaGlyThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 901 AAGTATTAAAGCTTGAAGTACCGTAAAGAGCGTTCATTCGGTGCATTTGCACA 960
Qy 505 SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
Db 961 CACGGA-----CAGAAATAAATGCGACGATTGTCGGGAA 996
Qy 525 AsnGlySerVal-----LeuProGluArgAlaLeuLeuValGlySer----- 538
Db 997 AATGGTCAGGTTTATCTGACTGGACTTCCACAGTCAGGGCAATTACAGGTTTCATGGGCG 1056
Qy 539 -----LeuGluTyrGlnLeuPro 544
Db 1057 AAAGATAAAACTCAAACTGTATTGTGCGAGTACAAGCTTCCT 1098

RESULT 22
US-09-974-300-2362
: Sequence 2362, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: TITLE OF INVENTION: Methods For Monitoring Multiple Gene
: FILE REFERENCE: 10085.500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: PRIOR FILING DATE: 2001-10-05
: PRIOR FILING DATE: 2000-10-06
: PRIOR FILING DATE: 2000-10-06
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2362
: LENGTH: 2055
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-2362

Alignment Scores:
Pred. No.: 0.0452 Length: 2055
Score: 104.00 Matches: 119
Percent Similarity: 36.26% Conservative: 100
Best Local Similarity: 19.70% Mismatches: 195
Query Match: 3.28% Indels: 190
DB: 10 Gaps: 36

US-09-857-669-2 (1-609) x US-09-974-300-2362 (1-2055)
Qy 20 TyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsn 39
Db 79 TATGCGCGCGGACACA---GTTGAAGACCAAAAGGCAACGGCT-----AAT 120
Qy 40 LysSerProAspThrGluSerValLysLeuLysProLysPheProValArgIleAspThr 59
Db 121 AAAACAGGAGCGGAGAGAG---GAAGTTCCCGCCAGCATACACTTTGAAAAAGAGATAC 177
Qy 60 GlnAspSerGluIleLysAspMetValGluGluHisLeuProLeuIleThrGlnGln 79
Db 178 TCTGACAACAGGGTC---GACCTTGCTCATCGATTTCGACATCATTAAGAAGGTGAAA 234
Qy 80 GluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnVal 99
Db 235 AAAGAAAGCGCAAGAAAGAGAAAAAGACAAAGCAGCGCCCATCTGAAAAAGACCATC 294
Qy 100 LysThrMetLeuArgSerLysGlyTyrPheSerSerLysVal-----SerLeuThrGlu 117
Db 295 AAATCTGTTT-----GAGAAAGCTGACTTCCGATTCGAGGACTCGCTGTGCACAG 345
Qy 118 LysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137

```

```

Db 346 GATTCACTCAACACCTCTCTGAGCGCCAGCCAGGAGAGCTTTCCCATCAGCGAGATTTCG 405
Qy 138 GlyValAlaIleLeuGlyAspIleLeuSerAsp-----GlyAsnLeuAlaGlu 153
Db 406 GTCATCACCCTCTCAATGACGTGATGAGCAAGAAATCCGCGCAGGAAAGCTCGATGAA 465
Qy 154 TyrTyrArgAsnAla-----LeuGluAsn 161
Db 466 GCCAAGACATGTAAACAAAGAACTGAAGGGCAATTCGATACCTTCAAAATTTAAACAC 525
Qy 162 TrpGlnGlnProValGlySer-----AspPheAspGlnAspSer 174
Db 526 GCGCGCTGATGAATCGGACGCTTTCGATCATTCGCCAATACGTATACGATCTCGACAAG 585
Qy 175 TrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAla 194
Db 586 ACGGACGAGAAGCGTCAAGAACCGCGCTCAGACGGTA----- 621
Qy 195 LysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnVal 214
Db 622 -----CAGCAGGTACAGATCAAGCAGGGCGAGATACTCGTTGAGGAAATCAG 669
Qy 215 ValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArg 234
Db 670 CTGATCGACAGGAA-----CTCTATCGAAGCTCGAATTTGACAGGG----- 711
Qy 235 TyrProGluGlnIleValSerGlyLeuAlaArgPheGlnPro-----GlyThrProTyr 252
Db 712 -----CTCTTGAGCGGAGCCCAACATCTTCAAACTGTGCGCGGACTG----- 753
Qy 253 AspLeuAspLeuLeuAspPheGlnAlaLeuGluGlnAsnGlyHisTyrSerGly 272
Db 754 ---CTGCTTTTAAATCGCGCTTTTCATCTCAGCTTTGGTT-----TACTAT----- 795
Qy 273 AlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSer 292
Db 796 -----TTCCAAAGCAATACAGACGCGGATTCGCGAAACAAAGTCG 837
Qy 293 ValThrGluValLysArgHisLysLysLeuGluThrGlyIleArgLeuAspSerGluTyrGly 312
Db 838 -----CTCATGCTGCTTTTCGTTCATCAGAGGATCGTTTA----- 873
Qy 313 LeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLys-----GlyTyr 329
Db 874 -----ATCGTTCATGGAAGTGCATCAGCCCTGTTTCAGCAGCTTGAATCGGCCAT 921
Qy 330 IleGlySerValValTrpAspMetAspLysTyrGluThrThrLeuAlaGlyIleSer 349
Db 922 ATCGGC-----TATCTCTGCGGATTGCGATTCGCGCGTGATG 957
Qy 350 GlnProArg-----AsnTyrArgGlyAsnTyrTrpThrSer----- 361
Db 958 CTCATTAAACTGCTCATCAATGAAGCGTCGCGATCCTGCGAGCATGATCTTTCCATT 1017
Qy 362 -----AsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
Db 1018 TCGCGAAGCATGATGTTTAACCGAGGCTGTCAGACGACGCTTCAAC-----TATGCA 1068
Qy 380 GlyGlyIleThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
Db 1069 ATCGGAACGATATTTCTTA-----ATCAGCTCGATGCGGGGTATACTGTTT 1113
Qy 400 LeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
Db 1114 TTAGGGAAG-----CATRACGCGAGGCGGAAA 1140
Qy 420 MetLeuThrAlaSerTrpLysArgGlnLeuAsnAsnValLeuHisProGluAsnGly 439
Db 1141 ATTCTCCAGCGCGTTTATTGCTCTCTGATCAACATCGTGTGTT----- 1185
Qy 440 HisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459

```

```

Db 1186 -----CTGTTTACCCTGCAGCTG 1203
QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
Db 1204 ATTCAAAACACGGCGCAACCGGGCTT-----GAGATGGGTACA 1242
QY 480 PheIleIleArgGlyGlnAla-----GlyTyrThrValAlaArgAspAsnAlaAsp 496
Db 1243 TATCTTGTTCATGGGGGGGTCTCCGGCATCGATCGTCATG----- 1284
QY 497 ValProSerGlyLeuMet-----PheArgSerGly---GlyAlaSerSerValArgGly 513
Db 1285 CWTGTTCGCGACTCATCGCGCTTTTGAAGCGGATTTGGCATCCTGCTCCACG----- 1338
QY 514 TyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluAla 533
Db 1339 -----ATGAAGCTCTCGAGCTTTCAAATCCGAATCATCCGCTGCTC-----CGCAA 1386
QY 534 LeuLeuVal-----GlySerLeuGlyTyrGlnLeuProPheThrArgThrLeuSer 550
Db 1387 ATTCTGACGGAAACACCGGCACATACCATCACAGCGTT---ATGTCGCGCAACTTATCT 1443
QY 551 -----GlyAlaValPhe 554
Db 1444 GAGCGGCATCGGAAGCGTCCGAGCAACCGGCTGCTGCGCAAGGTCGCGCGCTTATTAC 1503
QY 555 HisAspMetGly 558
Db 1504 CATGATATCGGC 1515

RESULT 23
US-09-974-300-2632
: Sequence 2632, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: TITLE OF INVENTION: Methods For Monitoring Multiple Gene
: TITLE OF INVENTION: Expression
: FILE REFERENCE: 10085.500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: PRIOR FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: 09/680,598
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/279,526
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2632
: LENGTH: 4557
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-2632

Alignment Scores:
Pred. No.: 0.176 Length: 4557
Score: 103.50 Matches: 130
Percent Similarity: 31.64% Conservative: 94
Best Local Similarity: 18.36% Mismatches: 222
Query Match: 3.26% Indels: 262
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-974-300-2632 (1-4557)
QY 25 AspLeuSerGluAsnLysAlaGlyPheAlaLeuPheLysAsnLysSerProAspThr 44
Db 571 GAAGTGAAGAGAGATCCAAATGTCAAAGGNCGTGTTGTCACACCCAGCTTCAACGCA 530
QY 45 GluSerValLys-----LeuLysProLysPheProValArgIleAspThrGln 60
Db 631 CAAAAAATAAAGGCTCCAGCAGCTCAAAAGATTCTTTCCGGTGAATTCAGACACTCG 690
QY 61 AspSerGluIleLysAspMetValGluGlnHisLeuProLeuIleThrGlnGlnGlu 80

```

```

Db 691 GAAGTAAACGGTT----- 702
QY 81 GluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLys 100
Db 702 ----- 702
QY 101 ThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr---GluLysAsp 119
Db 703 -----TATTTCAAACCAAAAGATAGCAAAACCATTTGATAATCA 741
QY 120 GlyAlaTyrThrValHisIleThrProGlyProArgThr----- 132
Db 742 GGAACCTTTCGATAAAGGCATCAACCCCGGCGAGTCAATGGAGCAGTCGACCTCAACAAA 801
QY 133 -----LysIleAlaAsnVal----- 137
Db 802 AAATTGGATCAGGTCAAAAATGCCAAGCTGCAGAACTTTCACAGAGGGGTGACCTATCGT 861
QY 138 -----GlyValAla 140
Db 862 TCTGTCAAAGTTTATCAGCTGGATGTCAACATAGATGTTTCAGTCAGCCGGCGAGATCAA 921
QY 141 IleLeu-----GlyAspIleLeuSerAspGlyAsnLeuAla 152
Db 922 GTTCTTTTCAAGGTACACCGTAGACGAGATGGAAATGTACATTCGACGCTGAGATGAT 981
QY 153 GluTyrTyrArg----- 156
Db 982 TCGGGCTTACCGCTCATCTATGTGACGGACATCGATGACTCCGCAAAACCGAAGAGC 1041
QY 157 -----AsnAlaLeuGluAsnTrpGlnProValGlySerAspPheAspGlnAspSer 174
Db 1042 GGAACCGCTGCATTTAGGAACAAGCAACGTTCCGGCGGACAAATCTGGAACCGGCTCT 1101
QY 175 TrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLys----- 189
Db 1102 GCAGAGCGCTCTGTCAACGGCAAAATACGGGAAATGATGAAAAATCGTCGACAGGTAT 1161
QY 190 -----GlyTyrProLeuAlaLysLeu 196
Db 1162 AAAGCGATTTCACAGCGTTCAGCTGGACGATTTTATTAATATGGAGAGAAAAATC 1221
QY 197 GlyAsnThrArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValVal 216
Db 1222 GACGAATCCAAAGCGAGCATAACAGACTCGCTTGTGTTCCGCTGATTTGCTTGTGTTCA 1281
QY 217 AspSer----- 220
Db 1282 GATTCCTTAAAGTGATTCCGCTTACTTTCATCAGAACCGCAGTCAGCAGCTGGAACG 1341
QY 221 ProIleAlaPheGly---AspPheGluIle-----ThrGlyThrGlnArgTyr 235
Db 1342 CCTCTAACCGAAGCGAAGGATATACGCTCTCTGATATGAACCGGCTTTGAAATCAAA 1401
QY 236 ProGluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAsp 255
Db 1402 TTTAACCAAGATGTTACCGGAGCTTACAAAATCAGCTATCAACCGAGGTCAACACGGGA 1461
QY 256 LeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerVal 275
Db 1462 GTAATCATCGAT-----AAATCAACAACCTATACGAAACACTGCGAGTG 1503
QY 276 GlnAlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGlu 295
Db 1504 -----ACCGAACAAGAGAAATCAAAAGAAAGCTTCAGGTACAGCC 1542
QY 296 ValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGly 315
Db 1543 GTTCACAGCAATTTGATAAAGGCTACTCAAAACATTTGATTACAG----- 1587
QY 316 LysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTpp 335

```

Db 1588 AAGAAACGGTGGATGGACGATCGCAATCAACAAAAACATACACGATGAACAATTGG 1647
Qy 336 AspMet---AspLysTyrGluThr-----ThrLeuAlaAlaGly----- 347
Db 1648 AAGTTGGATGACAAATTCGAAAGCGGGGCTCGCGCTCTCGACGGATCATTCGCGCTT 1707
Qy 348 -----IleSerGlnProArgAsnTyrArg----- 355
Db 1708 CAGGATGTGACGAATACAAACCGTTGAAGAGCGCAAGACATATACATTTGATATAAAA 1767
Qy 356 -----GlyAsnTyrTrpThrSerAsn----- 362
Db 1768 CCGGATCATGAAGGTTTTCCTCGGAACATCATCGGAGATTATCCGACAAACGACGCCAG 1827
Qy 363 -----ValSerTyrAsnAArgSer----- 375
Db 1828 CTTAAGATCATGATACCAACGATATCAATGACCGCATTTTCCAAAGAAATGTCAAAAC 1887
Qy 376 ArgAlaPheSerGlyGlyIleTyrTrpValArgAspArg-AlaGlyIleAspAlaArgLe 395
Db 1888 ACGCGCAATCA-----ACGTGG-----ACTGATCACAGCAGCACTGAACGCAAAAAT 1935
Qy 395 u-GlyAlaGlu---PheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuG 414
Db 1936 AAGGAACGACCACTTTTACCGCATGCCAGATCCCA-----C 1977
Qy 414 LysSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuAsnValL 434
Db 1978 AACGGCTTCAACACGCGTTATACACGCTGTTTCAAAAGAAAT----- 2021
Qy 434 euHisProGluAsnGlyHisTyrLeu-AspGlyLysIleGlyThrThrLeuGlyThrPhe 453
Db 2022 -CACGTGGAAATCGGCATCACTATATATGAGAG----- 2055
Qy 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGlu 473
Db 2056 CCGTCAAAAAACCGTTATATTAAGATTTCTTAGCAGATGACCAATTT----- 2106
Qy 474 AsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAsp 493
Db 2107 -----GAAAAGGCTCGGTGTCGTCAG-----ACGTATTCAGTTAATAAGAC 2151
Qy 494 AsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGly 513
Db 2152 GGTTC-----ATCAGAGGAGGAGACATCTCGCCCGCTCCCAA 2190
Qy 514 TyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAla 533
Db 2191 TATGATGTGAAGAGCGCTCGGTGCTGC-----AATAACAAACTTTGACCGTCCATTTG 2244
Qy 534 LeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaVal 553
Db 2245 AAACGGACGATTCGTCCTGCTATCTTATGAAATTTAAMACATCGCTAAAGGACAGGTC 2304
Qy 554 Phe---HisAspMetGlyAspAlaAlaAlaAsnPheTysArgMetLysLeuLysHisGly 572
Db 2305 ATTAAGCAGCATCTTACACCAATAAGGCCAGGTAC-----CATAT 2346
Qy 573 SerGlyLeuGlyValArg 578
Db 2347 GCGGATATTCAGACGGT 2364

RESULT 24
US-09-070-927A-132
; Sequence 132, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 5480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 132:
US-09-070-927A-132
Alignment Scores:
Pred. No.: 0.233 Length: 5480
Score: 103.50 Matches: 145
Percent Similarity: 30.06% Conservative: 88
Best Local Similarity: 18.71% Mismatches: 233
Query Match: 3.26% Indels: 309
DB: 10 Gaps: 38
US-09-857-669-2 (1-509) x US-09-070-927A-132 (1-5480)
Qy 42 ProAspThrGluSerValLys---LeuLysProLysPheProValArgIleAspThrGln 60
Db 2514 CCAGATGTTGACAGCATCGATGTTTAAAGTCCAGCTATTTCATTTGATCAAAACACGACA 2573
Qy 61 Asp-----SerGluIleLysAspMetVal----- 68
Db 2574 AGTAAACCCAGCTCAACGGTGGGGCGGTCACAGAAATCAATGATTATTACGATG 2633
Qy 69 -----GluGluHisLeuProLeuIleThrGln 77
Db 2634 TTATTTGCTGGGTAGTGCATCCAATTTGCTCAACGATCATATTGAAATACCAGTCAG 2693
Qy 78 GluGlnGluGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAsp 97
Db 2694 TCYGTGTAACAGATGTTGATATA-----GTCCTAGAGCTACCAGAG 2735
Qy 98 AsnValLys-----ThrMetLeuArgSerLysGlyTyrPheSerSer 111
Db 2736 CGAACAAAAATTCAAATTTTGCCCTCTGTTGTCAGAAAAAAGGCCAACATAAAAA 2795
Qy 112 LysValSerLeuThrGluLysAspGly-----AlaTyrThrVal 124
Db 2796 GTGTTGAAATGATCCAGCTCAAGGTATGTCAGATGCGGTGCGATGTGAACCTAT 2855
Qy 125 HisIleThrProGlyPro-----ArgThrLysIleAlaAsnValGlyValAlaIle 141


```
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-28

Alignment Scores:
Pred. No.:      0.0245      Length:      1164
Score:          103.00      Matches:      68
Percent Similarity: 36.31%      Conservatives: 50
Best Local Similarity: 20.92%      Mismatches: 114
Query Match:      3.25%      Indels:      94
DB:              10      Gaps:      13

US-09-857-669-2 (1-609) x US-09-974-300-28 (1-1164)

Qy 91 LeuAlaGluAlaProAspAsnValLysThrMetLeu----- 103
Db 64 CTGGCAGAGAGGTGAACAATACGGCGCAATGCTCTCGTATACGGCGGAGGCAGC 123

Qy 104 ---ArgSerLysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
Db 124 ATTAAGCGAAGCGTTTATACGATCAAGTCATTTCAATCCTTGAAGGGCGGCG--- 180

Qy 123 ThrValHis-----llethrProGlyProArgThrLysIleAlaAsnValGly 138
Db 181 ACCGTCATGAAGTCCCGCGCGCTCGAACCGAATCGCGGTGTGGCACTGTGAATAAGGA 240

Qy 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAla 158
Db 241 GTTGGGATCTGCAAGAGAACAGCATATTGACTTTCTTTGGCAGTCGGCGGCGGAGCGTC 300

Qy 159 LeuGluAsnTrpGlnGln---ProValGlySerAspPheAspGlnAspSerTrpGluAsn 177
Db 301 ATTGATTGTACGAAAGCAATATGCTGCGGAGGAGAAATACGACGGCGATGCGTGGGAT--- 357

Qy 178 SerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro----- 192
Db 358 -----ATTGTGAGAAANAACATATTCGCGGTGATGCGCTCCCG 396

Qy 193 -----LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr 209
Db 397 TTTCGAAACAGCTTTTAACGTTAGCAGACACAGCGCTCTGAAATGAACACTCGGGATCTGTGATC 456

Qy 210 Val-----AspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
Db 457 ACAAAATTGGCAAAACCAATGAAATAACGGCTGGGAGCCGCTCGTATTCCTCAAATTT 516

Qy 228 GluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheGln 247
Db 517 TCAATT-----CTTGA-TCCGGTCAACAGGTTTACCGT 548

Qy 248 ProGlyThrProTyrAspLeu-AspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAs 267
Db 549 CCCGAAAGACACACAGATTACGGCATTCGCACATGATGTCCCACGCTGTTTGAGCAATA 608

Qy 267 nGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVa 287
Db 609 TTTTCACCATACC-----GAAATAC 629

Qy 287 lProValLysValSerValThrGlu---ValLysArgHisLysLeuGluThrGlyIleAr 306
Db 630 CCCTTATCAGGACCGGATGTGCGAATCCCTGCTTAAACCGGTAATTGAACAGCTCCTAA 689

Qy 306 gLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAs 326
Db 690 GCTC----- 693

Qy 326 nLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr----- 342
Db 342 ----- 342
```

```
Db 694 -----ATTGAAGACCTAGAAAACCTATGAGCTGCCTGAAACGAT 731
Qy 343 -----ThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTy 358
Db 732 TCTGTATACAGGACCACTTCCGCTGAACGCGCATGCTATCAATGGCGCGAGACTG 791
Qy 358 rTrpThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAlaPh 378
Db 792 GGCACCGCACATATCGAGCAGCGCTGTTTCAGCCGCTATACGATATT-----CCGCA 842

RESULT 26
US-09-938-842A-2606
; Sequence 2606, Application US/0938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2606
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2606

Alignment Scores:
Pred. No.:      0.0366      Length:      1389
Score:          102.50      Matches:      89
Percent Similarity: 33.41%      Conservatives: 57
Best Local Similarity: 20.37%      Mismatches: 164
Query Match:      3.23%      Indels:      127
DB:              9      Gaps:      19

US-09-857-669-2 (1-609) x US-09-938-842A-2606 (1-1389)

Qy 89 GlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyr 108
Db 58 GGTCTTAATTGAGGAGAGAACTGTGAATTTCTGCTCTTTG----- 102

Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrValHisIleThrPro 128
Db 103 -----GGGAAATCAGAAAAAGAT----- 120

Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 121 -----CCGTTAAGGATTGTAAGTGTGTGCTGTGCTGGGGAAGTGTTTTGCA--- 171

Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGln-----Gln 164
Db 172 GCACCTTCTTCAAGAAAGCTATGGAGGTTTCAGGAGTAAGTTTCAGATCAGGATATGAGA 231

Qy 165 ProValGlySerAspPheAspGlnAspSerTrpGlu-----AsnSer 178
Db 232 AGACCTGGGAGAGCTGTGTATAGAGAACTGCAGAACATTTGTTGAAGTGCATCAATTCA 291

Qy 179 LysThrSerValLeuGlyAlaValThrArgLys-----GlyTyrProLeuAla 194
Db 194 ----- 194
```

```
Db 292 AGGGAAGATATCTTGGAGGAGATTGATAAGACGCTGTGCTTATCTGAATATGTCGAGGCA 351
QY 195 LysLeuGlyAsnThrArgAlaAlaValasnPro----- 205
Db 352 AGGCTTGTGTAGTACGACACTCTATGCTGATGAGATATTGAAAGACGGGTTTTGTCTTAAC 411
QY 206 -----AspThrAlaThrValaspLeuasnValValaspSerGlyArgProileala 223
Db 412 ATGGTTGATACCGCGCTTCTCTCTTAAGTTGTGACAAATCTGCAAGAAGCTCTCTGG 471
QY 224 PheGlyAspPheGluIleThrGly-----ThrGlnArgTyrProGluGln 238
Db 472 GATCCTGATATGTTGTTAATGATGCTTCAACTGAAACACGCTGGAAGTGTGTAAGAG 531
QY 239 IleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuaspLeuLeuLeu 258
Db 532 ATT---ACTAAGTAGTTGGAAGAGAGAATAACGGTTCCGATATTATCTCTCTGCAAG 588
QY 259 AspPheGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAsp 278
Db 589 GGTATTGAAACTGCTCTTGAA----- 609
QY 279 PheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArg 298
Db 610 -----CCAGTTCACATATCATTAACCTCAACCAAGATG 642
QY 299 HisLysLeuGluThrGlyIleArgLeuAspSerGlyTyrGlyLeuGlyGly---LysIle 317
Db 643 ATTCAATCAAGCAATGCTGTGCGCATGTGACATGCTCTGTATCTTGGTGGACCAACATT 702
QY 318 AlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMet 337
Db 703 GCTGCTGAATTTAC-----ACAAGCAATATGCCAATGCTAGATCTGTGGAGCT 753
QY 338 AspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsn 357
Db 754 GCTAAATGGAGAACCCCATGAGTAACTTCTTAAGACAACTTCATTTTCATT----- 804
QY 358 TyrTrpThrSerAsnValSerTyrAsnArgSerThrGlnAsnLeuGluLysArgAla 377
Db 805 GTTTGGGCAATAGTATGCTTGTGACACATGAAGTAATGGAGGCTC---AAGAATGTC 861
QY 378 PheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAla 397
Db 862 TACGCCATTGGA-----GCTGGTATGTTAGCGCGCTCACTAAC 900
QY 398 Glu----- 398
Db 901 GAGAGCGCTACAAGCAAGTCGGTGTATTTGCTCATTTGTACATCTGAGATGATATTTATA 960
QY 399 -----PheLeuAlaGluGlyArgLysIleProGlySerAspIleaspLeuGlyAsn 415
Db 961 ACTCATTTACTAGCAGAGACCCCTGAGAAACTTTCAGGCGCT-----TTGCTAGCTGAC 1014
QY 416 SerHisAlaThrMetLeuThrAla-----SerTrpLysArgGlnLeuLeuAsnVal 433
Db 1015 ACTTATGTGACCTTATTAAAGAGCGTAATGCATGCTGACGGTCAATGCTG----- 1065
QY 434 LeuHisProGluAsnGlyHisTyrLeuAspGlyIleGlyThrThrLeuGlyThrPhe 453
Db 1066 -----GCAAGGGTGAATAAATAGACACATGGGTGATAGC 1101
QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1102 ATAAGCGGCAAGGAATGATTCAGGGTGTCTTCTGACGTCGAGGACATTTTAC 1152
```

RESULT 27

```
US-09-956-004-14
: Sequence 14, Application US/09956004
: Patent No. US20020072595A1
: GENERAL INFORMATION:
: APPLICANT: Patrick J. Dillon et al.
: TITLE OF INVENTION: Nucleotide Sequences of Escherichia coli Pathogenicity Islands
```

```
: FILE REFERENCE: PB324D1
: CURRENT APPLICATION NUMBER: US/09/956, 004
: CURRENT FILING DATE: 2001-09-20
: PRIOR APPLICATION NUMBER: 08/976, 259
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/061,953
: PRIOR FILING DATE: 1997-10-14
: PRIOR APPLICATION NUMBER: 60/031,626
: PRIOR FILING DATE: 1996-11-22
: NUMBER OF SEQ ID NOS: 142
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 22671
: TYPE: DNA
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (19750)..(19750)
: OTHER INFORMATION: n equals a, t, g, or c
: NAME/KEY: misc_feature
: LOCATION: (20174)..(20174)
: OTHER INFORMATION: n equals a, t, g, or c
US-09-956-004-14

Alignment Scores:
Pred. No.: 2.7 Length: 22671
Score: 102.50 Matches: 129
Percent Similarity: 33.76% Conservative: 80
Best Local Similarity: 20.84% Mismatches: 212
Query Match: 3.23% Indels: 199
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-956-004-14 (1-22671)
QY 9 LeuLeuProAlaLeu-----PhePhePheProHisAlaTyrAlaProAlaAla 24
Db 11323 CTCCCTCCCTCCCTTTGCACTTTCTTCATTTTCTCGCGCAGCCATGCTGTCTCCGGT 11382
QY 25 Asp-----LeuSerGluAsnLysAlaAlaGly 33
Db 11383 GACCCAGCTGCANTTCAGCACACAGCAGCTTCTTGATGATAAACACGCGCCAGCGT 11442
QY 34 PheAlaLeuPheLysAsn-----LysSerProAspThrGluSerVal 47
Db 11443 GATGGCTGGAGCGACGACTGGCCGCTGACCATCACGCCGCTCTCCGGAACGCTCTCCCGT 11502
QY 48 LysLeuLysProLysPheProVal-----ArgIleAsp 58
Db 11503 ACTGAAGCTCCCTGCTTTACGGTGTCTACGATTTGTCAGTGGGCCACCGACTGACG 11562
QY 59 ThrGlnAspSerGlu-----IleLysAspMetValGluHisLeuProLeuIleThr 76
Db 11563 TCTGCAGAAACCGACGACGACTGGTGGCACCCGCTGGTGAATCAGTGTCTGAATATC----- 11616
QY 77 GlnGlnGlnGluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaPro 96
Db 11617 -----ACGGGACTGACCGCG-----GTCAGG 11637
QY 97 AspAsnValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThr 116
Db 11638 GATGCGGTGACGAGCGGCTATATACGCCGGGATATATACACGCGCGGCTTTCTGACA 11697
QY 117 GluLysAsp---GlyAlaTyrThrValHisIleThr----- 127
Db 11698 GACGAGGACCTTTCAGGGGGCGTACTGCACATAACGCTATGGAAGCAGCGCTGCAGCAA 11757
QY 128 -----ProGlyPro 130
Db 11758 ATCCGGCGGGAAGCGCTGACCTTCTTCGCCGACCCCTGAAGATGTTTTCCCGGGAATG 11817
QY 131 ArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsn 150
Db 11818 GAGGGGAAGGTTCTCACTTCGCGG-----ATATTGACGAG-GGG--- 11855
```

QY 151 LeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPhe 170
Db 11856 ATGGAGCAGATTATTCGTCTCGGTACGAGCGCGGTACAGATTCAATA---TCGCCCGGT 11912
QY 171 AspGlnAspSerTrpGluAsnSerLysThrSerValLeuAlaValThrArgLysGly 190
Db 11913 GACCTGAGGATGG---TCGGTGTGACACTGACGCGCATTCGCGGAA 11957
QY 191 TyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
Db 11958 TGGCTGTTCACAGGAGCGGTGGGATCGAACAGCGGCGAGAAATACCGGTACGGG 12017
QY 211 AspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThr 230
Db 12018 CAGTTAAATGGTCTTCTCTTTAAATCCTCTGGGCTGGCTGACAACTGGTTTCTC 12077
QY 231 GlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg---PheGlnProGly 249
Db 12078 ACCGGGAGCAGGAGCGTACTTTTCGGTGTGCATCATCGCAGGAAATTTTCCGCCCGGT 12137
QY 250 Thr-----ProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGluGlnAsn 267
Db 12138 GTCACTGCTCGGTATGGCTAT---ACCCTGGTGATACACG----- 12176
QY 268 GlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp-----Arg 281
Db 12177 -----TATTATGAGTGACTACTCTCAGCACCATTGATAACCGGGCTGGCGGTGGCGT 12230
QY 282 LeuGlnGlyAspArgValProValLysValSerValThrGluVal-----LysArgHis 299
Db 12231 TCCAGCGGAGAGCTCAGACTACCGGCTGGACTGTCGATCTCTCTTCGTAACGGG 12290
QY 300 LysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyr 319
Db 12291 GACATGAGACAGCAGCTG-----ACCGAGGCTGTCAGCAGCGCATTTATTCAC 12338
QY 320 AsptTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLys 339
Db 12339 AATTATCTGGAT-----GATGTCTGCTTCAGGCGCAGCAGCGT 12377
QY 340 TyrGluThrLeuAlaAlaGlyLysSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
Db 12378 AAATCACTTCATTTCTGCGGCTGAATCACACACAGTTCTGGTGGTGTCCGA 12437
QY 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
Db 12438 ACACGAATCCCGTATTCACACGG----- 12461
QY 380 GlyGlyIleTyrTyrValArgAspAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
Db 12462 GGGATGCCCTGGTTC-----GGCCACAA--- 12485
QY 400 LeuAlaGluArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
Db 12486 AGCAGACCACGGGAAAGG-----GGAGACCTGCGCGTAAATCAGTTCCGGAATGG 12536
QY 420 MetLeuThrAlaSerTrpLysArgGlnLeuAsnAsnValLeuHisProGluAsnGly 439
Db 12537 TCGGTGAGTGCCAGTTTTCACGCGCGCGTCCAGCAGGCTGTGGTGG----- 12584
QY 440 HisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459
Db 12584 ----- 12584
QY 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGlyThr 479
Db 12585 ---CTGACAGCGCTTATGCCACAGTGG-----TCACGGGACCGT----- 12620
QY 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSer 499
Db 12621 -----CTTCATGGGTGGGAACAACTAGCCTC----- 12647

QY 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 12648 -----GGGGGTGAGATTTCAGTCTGGCTGGCTTTTAAGGAGCAGGTATATTC 12689
QY 520 GlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 12690 -----TCGGGTAAATACGCGCGTTATCTG-----CGAAATGAGCTGTCTCTCTG 12737
QY 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGly 558
Db 12738 ---TTCCTCCTGCCATATGTGGGACAGTCCGTGACGTGACTGCACCTGAGCGCGGC 12791
RESULT 28
US-09-917-800A-1505
; Sequence 1505, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1505
; LENGTH: 15231
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 x59601
US-09-917-800A-1505
Alignment Scores:
Pred. No.: 1.9 Length: 15231
Score: 101.50 Matches: 113
Percent Similarity: 34.53% Conservative: 80
Best Local Similarity: 20.21% Mismatches: 226
Query Match: 3.20% Indels: 140
DB: 1.0 Gaps: 28
US-09-857-669-2 (1-609) x US-09-917-800A-1505 (1-15231)
QY 64 IleLysAspMetValGluGluHis---LeuProLeuIleThrGlnGlnGluGluVal 82
Db 12574 ATCATTTGACCTGAGGAGGAGCAGCGCTGCTGTGGAAGTGGCTATTAAGCGTGGTCTC 12633
QY 83 LeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMet 102
Db 12634 TTTGATGAGGAGATGACGAGATCTCTGACTGACCCCTCAGATGAC----- 12678

```
QY 103 LeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyr 122
Db 12679 -----ACCAAGGCTTCTCGACCAAC-----ACGAGGAGAACCTCACATAC 12723
QY 123 -----ThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly--- 138
Db 12724 CTGCAGCTGATGGAGCGTGTATCACTGACCCACAGCTGGCTGTGTCTCTCGCGCTG 12783
QY 138 ----- 138
Db 12784 AGGAAAGACGGGAGCGGAGACGCTCTCCAAAGTCTCACTGGCGCAAGCGCGGTG 12843
QY 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAla 158
Db 12844 GTGATTGTGGACCTTGAGACGGGCAAGAGATGTCACTGTATGAGCGCTTACCCCAAGGCG 12903
QY 159 LeuGluAsnTyrGlnProValGlySerAspPheAspGlnAspSerTyrPheGlnAsnSer 178
Db 12904 CTCATAGACCAACGACATACCTGGAGTTGTTCAGACGAGGAGTGCAGTGGGAGAAATC 12963
QY 179 LysThrSerValLeuGlyAlaValThrArgLys----- 189
Db 12964 ACCATCTCTCTCGGAGCGGCTCGTCAAAATCTATGATCATGACCGCGCTCTGGCGCG 13023
QY 190 -----GlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnPro 205
Db 13024 CAGTAGACATTGGTGACCGCATCACCAGAAGCTCATGACCGCTCAGACCTGACACGAG 13083
QY 206 AspThrAla-----ThrValAspLeuAsnValValAsp-----SerGlyArgProIle 222
Db 13084 TACCGCGTGGCACACTTCTATCAGCGAGTTTGCAGACATGCTCTCAGGCAACGCTGGT 13143
QY 223 AlaPhe-----GlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleVal 240
Db 13144 GGCTTCGCTCCGCTCTCTCTCTGGGCTCATCTCTCTACCC-----ATCAGT 13197
QY 241 SerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPhe 260
Db 13198 TCTGCTGTCTCTAGACCCAG-----CTAGCCCTCGGTGG 13230
QY 261 GlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp 280
Db 13231 TCTGATCTCTAGGAGACTGGCCAGCTGGCCGGCATC-----CTAGACACAGAG 13281
QY 281 ArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLys 300
Db 13282 ACTCTGGAG-----AAGGTGTCTCATCAGAGCGCCATGACCGCGCAAC 13323
QY 301 LeuGlu-----ThrGlyIleArg---LeuAspSerGluTyrGlyLeuGlyLys 316
Db 13324 CTGGTAGACAACATCACTGGCCAGCGGTGTGTGGAGCAGAGCGCTGCACCGGGGCATC 13383
QY 317 Ile-----AlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIle 330
Db 13384 ATTGACCCAGACACTGGTGAGCGCTTCCGGTCACTGAGGCTGTCAACAGGCGCTGGTG 13443
QY 331 GlySerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGln 350
Db 13444 GACAAGATCATGGTAGACCGTATCAATCTGGCCAGAGCGCTTCTGTGGGTTTGGAGAC 13503
QY 351 ProArgAsn-----TyrArgGlyAsnTyrTyrThr 360
Db 13504 CCAGGACCAAGACCAAGATGTCAGCTGCCAGCGCCCTGGAAGAGGC-----TGCGTT 13557
QY 361 SerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGly 380
Db 13558 -----TACTACGAGCGAGCGGCTTCTCTCAGGTGTCAGTACCTGACGGGT 13605
QY 381 GlyIleTyr-----TyrValArgAspArgAla 389
Db 13606 GGTCTGATTGAGCCCTGACACACCTGGCGGTGTGTCTCTGATGAAGCCTTGCAAGCTGGC 13665
QY 390 GlyIleAspAlaArgLeuGlyAlaGluPhe-----LeuAlaGluGlyArg 404
```

```
Db 13666 ACTGTGGATGCGCGCAGACCCAGACGCTCGTGATGTCTAGTGCCTACTTCCAAGTACCTC 13725
QY 405 LysIleProGlySerAspIleAspLeuGlyAsnSerHisAla-----ThrMetLeu 421
Db 13726 AGTGTGCCCAACAGCAACAGCTCAAGATCTCTTACAAGGACGCTCTGGATCGGAGCATGTG 13785
QY 422 ThrAlaSerTyrLysArgGlnLeuAsnValLeuHisProGluAsnGlyHisTyr 441
Db 13785 GAGGAGGCGACAGGCGCTGAGCTCTGTGAAGCGCGCACAGTCCAGCAAGGCTACTAC 13845
QY 442 LeuAspGlyLysIle-----GlyThrThrLeuGlyThrPheLeuSerSerThrAla 458
Db 13846 AGCCCGTCAGTGTGTCAGCGGCTCTGGCTACTGCTGTGTTCACGCACTGGTTCA 13899
QY 459 LeuIleAlaGlyThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLysLeuGly 478
Db 13900 -----CGACCGGCTCGAGCGCGCGC-----TCCGCTCGTGGC 13932
QY 479 ThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaAlaArgAspAlaAspValPro 498
Db 13933 AGCTTTGATGCACTGGC---TCTGGCTTCTCCATGACCTTTTCTTCTCTCTACTCT 13989
QY 499 Ser-----GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGlu 515
Db 13990 TCTCAGGCTATGGCGCGCGCTATGCCCTCAGGCGCTTACGCTCTCTGGGGCGCTGAG 14049
QY 516 LeuAspSerIleGlyLeuAla-GlyProAsnGlySerValLeuProGluArgAla 533
Db 14050 -----TCTGCAGTGGCTGATCCCGCAGCTGTATCTCTGCTTCCGCTCT 14095

RESULT 29
US-09-841-132-170
; Sequence 170, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 170
; LENGTH: 2949
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-841-132-170

Alignment Scores:
Pred. No.: 0.173 Length: 2949
Score: 101.00 Matches: 133
Percent Similarity: 32.85% Conservative: 70
Best Local Similarity: 21.52% Mismatches: 237
Query Match: 3.18% Indels: 178
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-841-132-170 (1-2949)
QY 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
Db 1003 TTCTTTAGTCAATGTAGCTGCGGAAAGGGGAGCTATTTATGCCAAAAGCTCTCG 1062
QY 128 Pro-----GlyPro-----ArgThrLysIleAlaAsnValGlyValAlaIle 141
Db 1063 GTTGCTAACTGTGGCCCTGTACAATTTTAAAGGAATATCGCTAATGATGCGGCGATT 1122
QY 142 -----LeuGlyAspIleLeuSerAsp 148
Db 1123 TATTTAGGAGAAATCTGGAGAGCTCAGTTTATCTCTGCTATTGAGATATATTTCGAT 1182
```

```
Oy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
|||||
Db 1183 GGGATCTTAAA-----AGACAGCCAAAGAGAAATGCTGCCGATGTTAATGGCGTA 1233

Oy 169 AspPheAspGlnAspSer-----TrpGluAsnSerLysThrSerValLeuGlyAla 185
|||
Db 1234 ACTGTGCTCTCAACAGCCATTTTCGATGGGATCGGGAGGAAATAACGACATTAAAGAGCT 1293

Oy 186 ValThrArgLysGlyTyrProLeu-----AlaLysLeuGlyAsn 198
|||
Db 1294 -----ANACAGGGATCAGATTCCTTTAATGATCCATCGAGATGGCAACGGAAAT 1347

Oy 199 ThrArgAlaAla-----ValAsnPro----- 205
|||
Db 1348 AACAGCCAGCGAGCTTCCAAACTTCTAATAAATTAACGATGTTGAAGGATACACAGGG 1407

Oy 206 -----AspThrAlaThrValAspLeuAsnValValAspSerGly 219
|||
Db 1408 GATATGTTTCTTAATCGAAGCAGTCTTGTACCAAAATGTTACGATAGACGAGGA 1467

Oy 220 Arg-----ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
|||
Db 1468 AGGATTGTTCTTCTGTAAGGCAAAATATCATCAGTGAATTCCTAAGTCAGACAGGTGG 1527

Oy 233 GlnArgTyrProGlu-----GlnIleValSerGlyLeuAlaArgPheGln 247
|||
Db 1528 ACTCTGATATGGAAGCTGGGAGTACATGATTTTGTAACTCCACACCACCAACAG 1587

Oy 248 ProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnGlnAlaLeuGlu----- 265
|||
Db 1588 CCTCCTGCCGCTAATCAGTTGATCAGCTTTCCAACTGCAATTTGCTCTCTTCTTTG 1647

Oy 266 -----GlnAsnGlyHis--- 269
|||
Db 1648 TTAGCAACAATCCAGTTACGAATCCTCTACCACATCTCCAGCGCAAGATTCATCCT 1707

Oy 270 -----TyrSerGlyAlaSerValGlnAla 277
|||
Db 1708 GCAGTCATTGGTAGCAAACTGCTGGTCTGTTTACAATTAAGTGGGCTATCTTTTGAG 1767

Oy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
|||
Db 1768 GATTTGATGATACACCTTATCATAGTATGATGTTGGTAGTCTTATCAAAAAATCAAT 1827

Oy 298 ArgHisLysLeuGluThrGly-----IleArgLeu 307
|||
Db 1828 GTCTCAAAATTACAGTTAGGAGTAAAGCCCAAGCTAATGCCCATCAGATTTGACTCTA 1887

Oy 308 AspSerGlu-----TyrGlyLeuGly-----LysIleAlaTyrAspTyrTyr 322
|||
Db 1888 GGAATGAGATGCTAAGTATGGCTATCAAGGAAGCTGAGAGCTTGGCTGGGATCCTAAT 1947

Oy 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr 342
|||
Db 1948 ACAGCAAAATAGTCTTATCTGAAAGCTACATGAGCTAAAGCTGGGTATATCCT 2007

Oy 343 ThrLeuAlaAleGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsn 362
|||
Db 2008 GGGCTGAGCGAGTAGCTTCTTTGGTTCCAAAT-----AGTTATGGGATCCATT 2058

Oy 363 Val-----SerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
|||
Db 2059 TTAGATATACGATCTGCGCATTCAGCAATTAACAGCAAGTGTGGAGCGGCTCTTATGTT 2118

Oy 380 GlyGlyIleTrp-----TyrValArgAspArgAlaGlyIleAsp 392
|||
Db 2119 CGAGGATATATGTTCTGGAGTTTCGAATTTCTTCTATCATGACGCGGATGCTTTAGCT 2178

Oy 393 AlaArgLeuGlyAlaCluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
|||
Db 2179 CAG-----GGATATCGGTATATTAGTGGGGTTATTCCTTAA---GGAGCAAACTCCTAC 2229

Oy 413 LeuLysAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
|||
Db 2230 TTTGGATCATCGATGTTTGGTCTAGCATTTACCGAAGPATTTGGTAGATCTAAAGATAT 2289

Oy 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle-----Gly 447
|||
Db 2290 GTAGTGTCGTTCCCAATCATCATCTGCTGCATAGGATCCGTTTATCTATACCCAA 2349

Oy 448 ThrThrLeuGlyThrPheLeuSerSerThrAlaLeuLeuLeuArgThrSerAlaArgAlaGly 467
|||
Db 2350 GCTTTATGTCGATCCCTATTTGTTGCGAGATGCGTTTATC-----CGTGTAGC 2397

Oy 468 TyrPhePheThrProGluAsnLysLeuGlyThrPheIleIleArgGlyGlnAlaGly 487
|||
Db 2398 TACGGGTTTGGGAATCAGCATATGAAA-----ACCTCA 2430

Oy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
|||
Db 2431 TATACATTTCCAGAGAG----- 2448

Oy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
|||
Db 2449 ---AGCGATGTTGCT-----TGGGATAATAACTGCTGCTGCGAGAGATTGGAGCG 2496

Oy 528 ValLeuPro-----GluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeu 543
|||
Db 2497 GGATTTACCGATTGCTGATTAATCCATCTAAGCTCTATTGGAATGAGTTGCGT----- 2547

Oy 544 ProPheThrArgThrLeuSerGlyAlaValPheHisAspMet-----GlyAsp 559
|||
Db 2548 CCTTCTGCTCAAGCTGAGTTTCTTATCCGATCATGAATCTTTTACAGAGGAAGCGAT 2607

Oy 560 AlaAlaAlaAsnPheLysArgMetLysLysLysHisGlySer---GlyLeuGlyValArg 578
|||
Db 2608 CAAGCTGGGCAATTCAGAGCGGACATCTCTAATAATCATCATGTTCTGTTGGAGTGAAG 2667

Oy 579 Trp-----PheSerProLeuAlaPro-----PheSerPheAspIleAlaTyr 592
|||
Db 2668 TTTGATCGATGTTTCTAGTACACATCTCTAATAATATAGCTTTATGGCGGCTTAT 2721

RESULT 30
US-09-841-132-182
: Sequence 182, Application US/09841132
: Patent No. US20020061848A1
: GENERAL INFORMATION:
: APPLICANT: Bhatia, Ajay
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: FILE REFERENCE: 210121.469C8
: CURRENT APPLICATION NUMBER: US/09/841.132
: CURRENT FILING DATE: 2001-04-23
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0
: SEQ ID NO 182
: LENGTH: 3021
: TYPE: DNA
: ORGANISM: Chlamydia
US-09-841-132-182

Alignment Scores:
Pred. No.: 0.179 Length: 3021
Score: 101.00 Matches: 133
Percent Similarity: 32.85% Conservative: 70
Best Local Similarity: 21.52% Mismatches: 237
Query Match: 3.18% Indels: 178
DB: 10 Gaps: 32

US-09-857-669-2 (1-609) x US-09-841-132-182 (1-3021)

Oy 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
:::||||| |||
::: |||
```

Db 1075 TTCTTTAGTAGCAATGTAGCTGCTGGAAAGGGAGCTATTATTCACAAAAGCTCTCG 1134
Qy 128 Pro-----GlyPro-----ArgThrIleAlaAsnValGlyValAlaIle 141
Db 1135 GTTGCTAACTGGGCCCTGTACAATTTTAAAGAAATACGCTAATGATGGTGGACGAAT 1194
Qy 142 -----LeuGlyAspIleLeuSerAsp 148
Db 1195 TATTTAGGAGATCTGGAGAGCTCAGTTTATCTGCTGATTATGGAGATATTTTCGAT 1254
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
Db 1255 GGAATCTTAA-----AGAACAGCCAAAGAGAGATCTCCCGATGTTAATGGCGTA 1305
Qy 169 AspPheAspGlnAspSer-----TrpGluAsnSerLysThrSerValLeuGlyAla 185
Db 1306 ACTGTCCTCTCAACAGCCATTTTCGATGGATCGGGAGGAAATAACGACATTAAAGAGCT 1365
Qy 186 ValThrArgLysGlyTyrProLeu-----AlaLysLeuGlyAsn 198
Db 1366 -----AAGCAGGCGATCAGATTCTCTTAAATGATCCCATCGAGATGGCAACCGAAAT 1419
Qy 199 ThrArgAlaAla-----ValAsnPro----- 205
Db 1420 AACCAGCCAGCCAGCTCTTCCAAACTTCTAAAAATTAAACGATGGTCAAGGATACACAGG 1479
Qy 206 -----AspThrAlaThrValAspLeuAsnValValValAspSerGly 219
Db 1480 GATATGTTTTTGTGTAATGGAAGCAGTACTTTGTACAAATGTTACGATAGCAGCAAGA 1539
Qy 220 Arg-----ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 1540 AGGATGTTCTTCGTGAAAGGCAAAATTTATCAGTGAATTCCTAAAGTCAGACAGTGGG 1599
Qy 233 GlnArgTyrProGlu-----GlnIleValSerGlyLeuAlaArgPheGln 247
Db 1600 AGTCTGATATGGAAGCGGGAGACATTTGGATTTGTAACATCCACCAACCAACCAACAG 1659
Qy 248 ProGlyThrProTyrAspLeuAspLeuAspPheGlnGlnAlaLeuGlu----- 265
Db 1660 CCTCCTGCGGCTAAATCAGTTGATCAGCGTTTCCAAATCTGCAATTTGCTCTTCTTTT 1719
Qy 266 -----GlnAsnGlyHls--- 269
Db 1720 TTAGCAACAATGCAAGTTACGATCCTCTACCAATCCTCCAGCGCAAGATTCATCCT 1779
Qy 270 -----TyrSerGlyAlaSerValGlnAla 277
Db 1780 GCAGTCATTGGTAGCACAACTGCTGTTCTGTACAAATTAGTGGGCTATCTTTTGTAG 1839
Qy 278 AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys 297
Db 1840 GATTGGATGATACAGCTTATGATAGGTATGATGTTGGCTAGGTCTTAATCAAAAATPCAT 1899
Qy 298 ArgHisLysLeuGluThrGly-----IleArgLeu 307
Db 1900 GTCCCTCAAAATACAGTTAGGCACTAAGCCCCAGCTAATGCCCCATCAGATTGACTCTA 1959
Qy 308 AspSerGlu-----TyrGlyLeuGlyGly-----LysIleAlaTyrAspTyr 322
Db 1960 GGAATGATGATCCCTAAGTATGGCTATCAAGGAAGCTGGAAGTTCGGTGGGATCCATA 2019
Qy 323 AsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThr 342
Db 2020 ACAGCAAAATAGTCCCTTATATCTGTAAGCTACATGGAGCTAAACCTGGGTATATCCT 2079
Qy 343 ThrLeuAlaAlaGlyLysSerGlnProArgAsnTyrArgLysAsnTyrTrpThrSerAsn 362
Db 2080 GGGCCTGAGCGGAGTAGCTTCTTGGTTCCTCAAT-----AGTTATGGGGATCCAA 2130
Qy 363 Val-----SerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
Db 2131 TTAGATATACGATCGCCGATTCAGCAATTCACAGCAAGTCTGGATGGCGCTCTTATTCT 2190

Qy 380 GlyGlyIleTrp-----TyrValArgAspArgAlaGlyIleAsp 392
Db 2191 CGAGGATTATGGGTTTCTGGAGTTTCGAAATTTCTCTATCATCAGCCGATGCTTTAGGT 2250
Qy 393 AlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAsp 412
Db 2251 CAG-----GGATATCGCTATATTAGTGGGGTTATTCTCTTA---GGAGCAAACTCCTAC 2301
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsn 432
Db 2302 TTTGGATCATCATGTTTGGTCTAGCATTTACCGAAGATATTGGTAGATCTAAAGATTAT 2361
Qy 433 ValLeuHisProGluAsnGlyHisTyrLeuAspGlyLysIle-----Gly 447
Db 2362 GTAGTGTGTGCTTCCAATCATCATGCTTGTCATAGTAGCCGTTTATCTATCTATCACCACAA 2421
Qy 448 ThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGly 467
Db 2422 GCTTTATGTGGATCCTATTGTTCCGAGATCGGTTATC-----CGTGTAGC 2469
Qy 468 TyrPhePheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAlaGly 487
Db 2470 TAGCGGTTTGGCAATCAGCATATGAAA-----ACCTCA 2502
Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
Db 2503 TATACATTTGCAGAGG----- 2520
Qy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
Db 2521 ---AGCGATGTTTCGT-----TGGATAATAACTGTCTGCTGGAGAGATTGGAGCG 2568
Qy 528 ValLeuPro-----GluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeu 543
Db 2569 GGATTACCCATTTGTGATTACTCCATCTTAAGCTCTATTGTAATGAGTTGCGT----- 2619
Qy 544 ProPheThrArgThrLeuSerGlyAlaValPheHisAspMet-----GlyAsp 559
Db 2620 CCTTTCTGCAAGCTGAGTTTCTTATGCGCATCATCAATCTTTTACAGAGGAAGCGAT 2679
Qy 560 AlaAlaAlaAspPheLysArgMetLysLeuLysHisGlySer---GlyLeuGlyValArg 578
Db 2680 CAAGCTCGGGCAATCAAGAGCGGACATCTCTAAATCTATCAGTTCTCTGTGGGTAAG 2739
Qy 579 Trp-----PheSerProLeuAlaPro-----PheSerPheAspIleAlaTyr 592
Db 2740 TTTGATCGATGTTCTAGTACACATCCTAATAAATATAGCTTTATGGCGGCTTAT 2793

RESULT 31

US-09-070-927A-527

; Sequence 527, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch

; Patrick J. Dillon

; Steven Barash

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 382

; CORRESPONDENCE ADDRESS:

; ADDRESS: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070.927A

```

: FILING DATE: 04-May-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/046,655
: FILING DATE: 1997-05-16
: APPLICATION NUMBER: 60/044,031
: FILING DATE: 1997-05-06
: APPLICATION NUMBER: 60/066,009
: FILING DATE: 1997-11-14
: ATTORNEY/AGENT INFORMATION:
: NAME: Kenley K. Hoover
: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 527:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7107 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 527:
US-09-070-927A-527

Alignment Scores:
Pred. No.: 0.67 Length: 7107
Score: 101.00 Matches: 126
Percent Similarity: 33.22% Conservative: 77
Best Local Similarity: 20.62% Mismatches: 224
Query Match: 3.18% Indels: 184
DB: 10 Gaps: 34

US-09-857-669-2 (1-609) x US-09-070-927A-527 (1-7107)
Qy 25 AspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThr 44
Db 2306 GATGTATCTATTAGACATANTACACCTTTTAAAGTAGTAAAGGAGAGCT-----ACA 2359
Qy 45 GluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIle 64
Db 2360 GAAGCTATCCGT-----AAAATAGATTCCTCATTCGATACATTTAAACACGAGATTA 2410
Qy 65 LysAspMetValGlu-----GluHisLeuProLeuIleThrGln 77
Db 2411 ACAGAAATGATTGAAAAAGAGATAGTAGTAATCCACAGGACCTATACTATTATTCGGA 2470
Qy 78 GlnGlnGluValLeuAsp---LysGlnGlnThrGlyPheLeuAlaGluGlu----Ala 95
Db 2471 GACAACTCTTATAGTGGATGCTTAAGAGATGTTGAAATTCGACGTAAACAAAATCCCAACC 2530
Qy 96 ProAspAsnVal-----LysThrMetLeuLeuArgSerLysGly 107
Db 2531 ACAGAAATATTAAATAAGCACGAGTAGTTGGTTCTAAATGTGCTTTGGGACAAAAC 2590
Qy 108 TyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaIleThrValHisIleThr 127
Db 2591 AAATATAATCCAGTCTTAATAGTCTCTGAAGAAGTAAACAGATATCTTCAACTAGCTTTA 2650
Qy 128 ProGlyProArgThrLys-----IleAlaAsnValGlyValAlaIle----- 141
Db 2651 AATGGTTTCATCAATGATGCTGCTTGATATTAAATAATACTCATCGCGCACTTCGTT 2710
Qy 142 ---LeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluIle----- 154
Db 2711 AAATTAATGACCGGATAAAGATGGGATACTTGTATAGATACGACGAACAAACATGGGAAT 2770
Qy 155 -----TyrArgAsnAla-----Leu 159
Db 2771 GGAAACCAACTACACAATGCTGATCTCTGATGGTATGGAAGAAAGTGACGATTTGGAAT 2830
Qy 160 GluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLys 179
Db 2831 TTTAATGATACTGATCCTTTTGGTGAGTCGCTAGCAT-----TGCCTTTGATATAAAT 2881
Qy 180 ThrSerValLeuGlyAlaValAlaThrArgLysGlyTyrProLeuAla----LysLeuGlyAsn 198
Db 2882 GGAGAAAAATAAATAATAGTAACAACCGGACTACTGATAGATAAGTGGAAGAAATAGGAAT 2941
Qy 199 ThrArgAlaAlaValAsnProAspThr-----AlaThrValAspLeuAsn 213
Db 2942 AAC-----AACTATGATGACTGAAATGTCTACCCCTAGAACCTGTTCAATTAATA 2989
Qy 214 ValValValAspSerGlyArgProIleAlaPhe----- 224
Db 2990 AAGCTTACTGATAACGGAGAAAAATCTGATTTCTGAAATTAGTCTTCTGAAAGATAGAAAA 3049
Qy 225 GlyAspPheGluIleThrGlyThr-----GlnArgTyrProGlu 237
Db 3050 GGGACTTTTGAATTTTACTGGATTGAGTGTGTAATTAACCAAGGGAGATAGACTCGTTGTT 3109
Qy 238 GlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeu 257
Db 3110 AAAATTATTACTAAT-----GAAGTACAAAGAGAA 3139
Qy 258 LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHis----- 269
Db 3140 ATTGATCGACAAAGAAATGTTGTTCAAGTTGGTTATGATAATCCAGAGGTATCTGAAGAA 3199
Qy 270 -----TyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly----- 284
Db 3200 GTAATTGTTCCAGGGCCGACAGTAACAGTCTCTTTTGACTTTTAAATTAATAACAACACTGAT 3259
Qy 285 -----AspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
Db 3260 CCTGACTTTTGACCGTCAACCACTT---ACTCAAGTAGTAGAATTAGAAAA----- 3307
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321
Db 3308 -----GTTTCAAGGGCTTTTAAATAGTATT 3331
Qy 322 TyrAsnLeuPheAsnLysGlyTyr-----IleGlySerValValTrpAspMetAspLys 339
Db 3332 TTTACTCCAGAACGAAAGGCTATAAATTTATTTGA-----TGGAATACTGATAAA 3382
Qy 340 TyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrp 359
Db 3383 -----CTTGGAGTAGCAGACAAACTGTCAAT----- 3406
Qy 360 ThrSerAsnValSerTyrAsnArgSerThrThr-----GlnAsnLeuGluLysArgAla 377
Db 3407 ---GATGGTCTAGTTTAAATGAAGATATAACGGCTATTGCTCAATGGGAAAAAGAACCT 3463
Qy 378 PheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAla 397
Db 3464 AACCTAGCTGGA-----GAAGTACATGCTCCACAGGCTCCT 3499
Qy 398 GluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsn---Ser 416
Db 3500 -----ATTGAAGAGGTAAGATATTCTCTAAATAATAGTAGTACAGTCCCAATAATCA 3553
Qy 417 HisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnAsnValLeuHisPro 436
Db 3554 GGTCTGTAAATGTAAGTGGAGAACTTCCAAAGGGCTCTTTGGGTTATCTATAAAT--- 3610
Qy 437 GluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSer 456
Db 3611 -----GAAACAGCTAATCTAATTTGTTACTCCACTTATCAAT 3646
Qy 457 ThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys 476
Db 3647 GACTGGATTGATGGAGAAATAGTCGG-----GAAGTAAAG 3682
Qy 477 LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAsp 496
Db 3683 ATTCCAGTTTACAATTAGTAATCGTGACGAAAGGTTATGCTG-----GAA 3727

```


Oy	497	ValProSerGlyLeuMetPheArgSerGlyAlaSerSerValArgGlyTyrGluLeu	516
		: :	
Db	3728	GTTCCATTAAAGATGCCGTGATCTCGACGGCAGCACGATAAAGCAGCAGAAA	3784
		: :	
Oy	517	AspSerIleGlyLeuAlaGly-----ProAsnGlySerVal	528
		:	
Db	3785	---GACGCAGCAAGAAGTCGACCATGAAGAACCGCAAGTGAAACCACGCGTTTAGTG	3841
Oy	529	LeuProGluArgAlaLeuValGlySerLeu	539
Db	3842	ACTCCAGAGGAAAAAGCCGACGTGATGCGCTTG	3874

RESULT 32

```

US-09-815-242-6700
; Sequence 6700. Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EUITRA.011A

```

Alignment Scores:		
Pred. No.:	0.184	Length: 2820
Score:	100.50	Matches: 140
Percent Similarity:	29.97%	Conservative: 89
Best Local Similarity:	18.42%	Mismatches: 248
Query Match:	3.17%	Indels: 287
DB:	10	Gaps: 36

US-09-857-669-2 (1-609) x US-09-815-242-6700 (1-2820)

Qy	42	ProAspThrGluSerValLys	---	LeuLysProLysPheProValArgIleAspThrGln	60
Db	217	CCAGATGTTGACAGCATCGATG	TTTAAGTCCGATATTTC	CAATGTGATCAAAAGACGACA	276
Qy	61	Asp-	-----	SerGluIleLysAspMetVal-	68
Db	277	AGTAAAAACCCGCGTCAACGGTGGG	GACGGTCAACGGAAATCAATGATT	ATTTACGATTG	336

Qy	352	ArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGln	371
Db	1231	---GGGTATTCGTTTGAATCCGCAAGCTTTAGCTGTAAAAATCAATGGCACGCACATTTGGT	1287
Qy	372	AsnLeuGluLysArgAlaPheSerGlyIleTrpTyr	384
Db	1288	GAAGTCAGTGAATTACCAATAAAAAATCGGTCCAATTTTGAAGGTGTGCTTTATCT	1347
Qy	385	-----ValArgAspArgAlaGlyIle	391
Db	1348	GACACAGAAATACAGATGCTCGGCCAATTTTAAAGAAGTCGAAGATCGGTTAACCTTC	1407
Qy	392	AspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIle	411
Db	1408	TTAAAAAATGTTGGGTACATTACTTAACTTAACTCGA-----GCTGCTGGGTACACTT	1461
Qy	412	AspLeuGlyAsnSerHis-----AlaThrMetLeuThrAlaSerTrpLysArg	427
Db	1462	TCGCGCGGAGACACAAAGAAATTCGCTAGCACACAAATTTGGATCAAAAC-----	1512
Qy	428	GlnLeuLeuAsnValLeuHis-----pro	436
Db	1513	-----TTATCAGCGCTCTTTTACATTTTAGATGAACCATCAATTTGTTTGCATCAACGA	1566
Qy	437	GluAsnGlyHisTyrLeuAspGly-----LysIleGlyThrThrLeu---	450
Db	1567	GATACGACCGCTTGATTGACTTCATTTGAAANAATATCGGGGACTTTAGGCAATACATTTAAT	1626
Qy	451	-----GlyThrPheLeuSerThrAlaLeuIleArgThrSerAla	464
Db	1627	GTGCTGGAACACGATGAAGATACAATGATGCTCGGATTACTTGTATTGATGGGGACCT	1686
Qy	465	ArgAlaGlyTyrPhe-----PheThrProGluAsnLysLysLeu	477
Db	1687	GGCGCTGGACATCTCGGGCGTGAAATTTGACCGCGTGTACCGCGGAAGAGTGTGCTAAA	1746
Qy	478	GlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspVal	497
Db	1747	AATCCGCAATTCATTGACTGCACAA-----TATCTTCTGGGAAAAAGTGATTCTGTGA	1800
Qy	498	ProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAsp	517
Db	1801	CCAAAGACGTCGCAAGGTAAATGGCAAGCGATCAAGTCACTGCTGCCAGTGAANAAT	1860
Qy	518	SerIle-----GlyLeu	521
Db	1861	AATTTAAAAAATGTCAGTGTGAATTTCCGTAGTGAATTTGTGCTGCACAGCGCTT	1920
Qy	522	AlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeu-----	535
Db	1921	TCAGGTTCAGGGAAGATCTTACTTCAATCAATTTTGAANAAGCACTGGCCCAAAA	1980
Qy	535	-----	535
Db	1981	TTAAATCGTAATTCCAAATAAACCTGGGAAACACAAAAGTATCACTGTTATGAAGCAATT	2040
Qy	536	-----ValGlySerLeuGluTyrGlnLeuProPhe	545
Db	2041	GAAAAGATTGTTGACATTCATCAAAAGTCCANTTGTGCGAAGCCCAAGCAAGTATCAGCG	2100
Qy	546	ThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAspAlaAlaAlaAsnPheLys	565
Db	2101	ACTTATACT-----AGTGTTTTGTATGACATTCGGATTTGTTGGCCCAACAAAT	2151
Qy	566	ArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaPro	585
Db	2152	GAAGCCCAAG-----GTTCTGTGTTATATAAAAA---GCCCGC	2184
Qy	586	PheSerPheAspIleAlaTyrClyHisSerAspLys-----Lys	598
Db	2185	TTTAGTTTAAACGTCAAAGGTGGCGGCTTGTGAAGCTTGTCCCGGGGATGTTATCATCAAA	2244
Qy	599	IleArgTrpHis	602

```

Db      2245  ATTGAATGCAC  2256      ||||
RESULT 33
US-09-070-927A-341/C
: Sequence 341, Application US/09070927A
: Patent No. US20020120116A1
: GENERAL INFORMATION:
: APPLICANT: Charles A. Kunsch
:           Patrick J. Dillon
:           Steven Barash
: TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
: NUMBER OF SEQUENCES: 982
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/070.927A
: FILING DATE: 04-May-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/046,655
: FILING DATE: 1997-05-16
: APPLICATION NUMBER: 60/044,031
: FILING DATE: 1997-05-06
: APPLICATION NUMBER: 60/066,009
: FILING DATE: 1997-11-14
: ATTORNEY/AGENT INFORMATION:
: NAME: Kenley K. Hoover
: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 341:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 13884 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 341:
US-09-070-927A-341

```

```
QY 88 ThrGlyPheLeuAlaGluGluAlaProAspAsnVallysthrMetLeuArgSerLysGly 107
|||||
Db 11708 GCGGCGCTTTTAAACCAATATTATACAGAT-----CCTGAACGCAAGAT 11664
QY 108 TyrPheSerSerLysValSerLysLeuThrGluLysAspGlyAlaTyrThrValHisIleThr 127
|||
Db 11663 GTCGCCACAGACATGAATTCCTAAGTGAGTCTTCGGGTCTATATTGGCTAATTAGCT 11604
QY 128 ProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSer 147
|||
Db 11603 -----TTTGATACG 11595
QY 148 AspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGly 167
|||
Db 11594 CAGGACGCTTTTCATAATTTTACAAGCAGAGAAAGACACCTTTTACGACGCGCTCCAA 11535
QY 168 SerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr 187
|||
Db 11534 TTTAGTATTGCAATTGACGAGCAGGCAATTAATAC----- 11499
QY 188 ArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThr 207
|||
Db 11498 -----AATGTCATCCCTCGATTGAT----- 11478
QY 208 AlaThrValAspLeuAsnValVal-----ValAspSerGlyArgProIleAlaPhe 224
|||||
Db 11477 -----GATTACGGATCATTCGTTCAATTAATGAACGAGG----- 11442
QY 225 GlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAla 244
|||
Db 11441 GGTCAATTTTAAA-----ACAGATCAGTATGATCAACAAATTAATAAATTTAGCAAG 11391
QY 245 ArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnAlaLeu 264
|||
Db 11390 AGTTTATGAAACCTCTATGAAGAT---AATATTCAATGATTT----- 11346
QY 265 GluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGly 284
|||
Db 11345 -----TAGCATAGTAATCAAAACAA----- 11325
QY 285 AspArgValProVallysaValSerValThrGluVallysaArgHisLysLeuGluThrGly 304
|||
Db 11324 -----CAGCTTTCAGAACTCTCTTTGTTTATATTGACTTAATACGCTG 11280
QY 305 IleArgLeuAspSerGluTyrGlyLeuGlyGlyLysIleAlaTyrAspTyrTyrAsnLeu 324
|||
Db 11279 GGTATCTATACAAAGATTTGGGATTTTCACACACTTTCAGTATCATCATTACCAATTA 11220
QY 325 PheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGluThrThrLeu 344
|||
Db 11219 ATAGATGATGGCTATATTCTGAT-----GATTACCTCTTTATCAACGAAAGTTT 11169
QY 345 AlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSerAsnValSer 364
|||
Db 11168 AAT-----CATCAACAAATAATAACGAAACAA-----GGAACACTTAATATTATT 11121
QY 365 TyrAsnArgSerThrThrGlnAsnLeuGluLys-----ArgAlaPheSerGlyGlyIle 382
|||
Db 11120 GAATCTTTTATTAACGATTCCTCATTTAAGTGAAGTGGGTATGCGCAGACAGACTAGTATT 11061
QY 383 TrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGlu 402
|||
Db 11060 GATTTTGTGCGAAACAA-----GTTCAGCAAGGAGCGCTTTT-----AAC 11019
QY 403 GlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThr 422
|||
Db 11018 TCGTATGATTAAATGATGCGCCGCTGGATAAAATCAATCTGCAGCTAGTTATGCGATA 10959
QY 423 AlaSer 424
|||
Db 10958 GCGGCT 10953
```

```
RESULT 34
US-09-070-927A-17
; Sequence 17, Application US/09070927A
; Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Parash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 382
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PH369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-070-927A-17

Alignment Scores:
Pred. No.: 8.05 Length: 32768
Score: 100.50 Matches: 102
Percent Similarity: 36.29% Conservative: 70
Best Local Similarity: 21.52% Mismatches: 175
Query Match: 3.17% Indels: 127
DB: 10 Gaps: 29

US-09-857-669-2 (1-609) x US-09-070-927A-17 (1-32768)
QY 17 PROHISALATYRALAProAlaAlaAspLeu-----SerGluAsnLysAla 31
|||
Db 536 CCAAAAGAGATTCTCAATGCAATTCGAATTCGGAATTGGAACCTTACTACGTGACCGAACCAAGCA 595
QY 32 Ala---GlyPheAla---LeuPheLysAsnLysSerProAspThrGluSerValLysLeu 49
|||
Db 596 TCTAATGGTCTCTGCAATACCTTTAAA-----GCAGTAAAGTC 634
QY 50 LysProLysPheProValArgIleAspThrGlnAspSerGluIleLysAspMetValGlu 69
|||
Db 635 GAATTGAAGTAT-----GCCAAT 652
QY 70 GluHisLeuProLeuIleThrGln-----GlnGlnGluGluValLeuAspLys 85
|||
```

Ds 653 CAACCGCTGGCCCTGTGACCAAGCAATATCAAAAGGCAAAACCAAGAAATTTACTGGG--- 709
Qy 86 GluGlnThrGlyPheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSer 105
Ds 710 ---GAAACAACTCTGATAAAGAGACAGGATATCGAGATAAACACAGGCGAGGCA 766
Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGly----- 120
Ds 767 GTCTTTGAGGGGACTAATATACCTTGTCTTACTGCTAAGGATGGAAGAGCGGTTAAATGG 826
Qy 121 -----AlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnValGly 138
Ds 827 AGTGAGCTTTTAAACCAAGATTTGGTAAAGGCAACCAAGCTTCTGATGAACAGTACT 886
Qy 139 ValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla----- 152
Ds 887 CTGGCT-----TTAGATGAAGAAGATCAAGCTGCGCTTTAAACACTTAGCC 931
Qy 153 -----GluTyrTyrArgAsnAlaLeuGluAsnTyrGlnGln-----ProValGly 167
Ds 932 ATCAACGAGTACTAC-----TGGCAAGAACCAACAGCACCTGAAGGA 973
Qy 168 SerAspPheAspGlnAspSerTrp-----GluAsnSerLys 179
Ds 974 TATACCTTGGATGAACGAGATATCTGTTCYATCAAAAAGTGTGATGAACGAAAAA 1033
Qy 180 ThrSerValLeu-----GlyAlaValThrArgLysGlyTyrPro 192
Ds 1034 AATCCGCTTATACGAGATGTCGAGATGTCGCGCAAGAACAGATATTTCGTTTGGGTTTCAT 1093
Qy 193 LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaValAspLeu 212
Ds 1094 TCTTCAAAATTCGCGATCGCGYGTGTC-----ACTGCGCAAACTGGGTTT 1141
Qy 213 AsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Ds 1142 AATGACTTAACTTTAAGTGTCGCCATTGAA---GGGCAACATGAGATCACAGTGCT 1198
Qy 233 Gln-----ArgTyrProGluGlnIle-----ValSerGlyLeuAlaArgPhe 246
Ds 1199 GAAGATRAACAAATCACAGCTTATATAGCAATTTAGGCTTTGATGCTCGGAAAGTTT 1258
Qy 247 GlnProGlyThrProTyrAspLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 266
Ds 1259 GAA---AATCTCTCTATGGG---GATTATTTACTTGAAGAAGTAGAGCGCCAGAGGA 1312
Qy 267 AsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArg 286
Ds 1313 TTTCAAAAATTAACGCCATTAGCAATTCGGTCTACTTTTAAGGAAAAACAAGAGACTTT 1372
Qy 287 ValProValLys-----ValSerValThrGluVal---LysArgHisLysLeuGluThr 303
Ds 1373 GTGAAGAGTGTAGTGTCTTACCATTACCGAAGAGCCCAAGCAACCAATTAACAG 1432
Qy 304 GlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsn 323
Ds 1433 -----GTGACCGTTCTCTACGAAAG 1453
Qy 324 LeuPheAsnLysGlyTyrIleGlySer-----ValValTyrAspMetAspLys 339
Ds 1454 CTGACGAACAAGCAATTTCTGTAGTTGAACCGCTTGTGATGCTCTATGATTTCCCTGAG 1513
Qy 340 TyrGlu-----ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Ds 1514 GAGAGAGATAGTTGACTTCTCTGCACTTGAAGAGAGCGGAATATAAAATTAATCAATPACC 1573
Qy 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Ds 1574 CTTGATTTTACCAGCTAGTGTGATAAATTTGAGATATACTTGTGATGAATCAAGAGAGAC 1633
Qy 377 AlaPheSerGlyIleTyrValArgAspArgAla---GlyIleAspAlaArgLeu 395
Ds 1634 -----TGGTATGTCTAGCTCAGGCCATTGATGTGGAAGCCACAAAA 1675

Qy 396 GlyAlaGluPheLeuAlaGluGlyArgLys-----IleProGlySerAspIleAsp 412
Ds 1676 GCTGCCCAAGAAAAAGCAAAAGCCAAACAGTGGTAAATTTGCCGAAACCAACGACCACT 1735
Qy 413 LeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLys 426
Ds 1736 TTGGCGAACCAAGAGAAA-----ACAGGTACTTTGGAAA 1768

RESULT 35

US-09-815-242-4584
; Sequence 4584, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 4584
; LENGTH: 6258
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4584

Alignment Scores:

Pred. No.:	0.716	Length:	6258
Score:	100.00	Matches:	103
Percent Similarity:	35.05%	Conservative:	81
Best Local Similarity:	19.62%	Mismatches:	215
Query Match:	3.15%	Indels:	126
DB:	10	Gaps:	25

US-09-857-669-2 (1-609) x US-09-815-242-4584 (1-6258)

Qy 80 GluCluValLeuAspLysGluGlnThrGlyPheLeuAlaGluAlaProAspAsn--- 98
Ds 2899 GAAATATTTTAAATTAATAATAGTGGATCAAAATTTAGCAACAAAGCTGCGTTGAAACGCG 2958
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGlu--- 117
Ds 2959 TTGTCACAAGTCACAAATGCGAAAGGTGCCCTAAATGTTAACCATTAATTTAGAGCAAGCT 3018
Qy 118 LysAspGlyAlaTyrThrVal-----HisIleThrProGlyProArgThr 132
Ds 3019 AATCAANTGCAACCACTACTATTAACGCACTTCAACATTTTAACCACTGCACAAAAGAT 3078
Qy 133 LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Ds 152

Qy	80	GIuclVValLeuAspLysgluclnThrGlyPheLeuAlaGIuGluAlaProAspAsn---	98
		: : : : : :	
Db	9730	GAaNTATTTAAATAAATAATAGTGGATCAAAATTTAGACAAGGTCGGTGTAAACACGG	9789
Qy	99	ValYusThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGIu---	117
		::: : : : : : :	
Db	9790	TTGTCAACAGTGACAAATGCGAAAGGTGCCCTAAATGGTAACCAATAATTTAGAGCAAGCT	9849
Qy	118	LysAspGIValATyThrVal-----HisIleThrProGlyProArgThr	132
		: : : :	
Db	9850	AAATCAATGCAACACACTACTATAACCGAGCTTCACATTTTACAACTGCACAAAAGAT	9909
Qy	133	LysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla	152
		: : : : : : :	
Db	9910	AAATG-----	9915
Qy	153	GIuTyTyArgAsnAlaLeuGIuAsnTrpGlnGlnProValGlySerAspPheAspGln	172
		: : : : : :	
Db	9916	-----AAACAACAGTGCACACGCCACAAAATGTCAGGT-----GTA	9954
Qy	173	AspSerTrpGIuAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyPro	192
		: : : : : : : :	
Db	9955	GATACGTGTTAAATCAAGTGCCACACATTAATGGTGCT-----	9993
Qy	193	LeuAlaLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThr-----	209
		: : : : : : : :	
Db	9994	-----ATGGGTACGTAAAGAAATGACTTCACATACACAGCTAGGAATAATGCC	10044
Qy	210	-----ValAspLeuAsnValValValAspSer	218
		: : : : : : :	
Db	10045	CAAACTATCTTGATGCTACAGAAAGTAACAAAACAACTATAACAACTGCTGTATGAT	10104
Qy	219	GlyArgProIleAlaPheGIAspPheGluIleThrGlyThrGlnArgTyProGIuGln	238
		:	
Db	10105	-----GCTAAATGGTGCTATTAAATGCAACAGCAATCCAAATATGGATGCTAAT	10152
Qy	239	IleValSerGlyLeuAlaArgPheGlnProGlyThrProTyAspLeuAspLeuLeu	258
		: : : : : : : :	
Db	10153	GCAATTTACCAATCGCTACACAGCTGACATCAACAGAAANTGCATAGATGGTACACAT	10212
Qy	259	AspPheGlnGlnAlaLeuGlu-----GlnAsnGlyHisTySerGlyAlaSerValGlnAla	277
		::: : : : : :	
Db	10213	AAATTTAACCGAAGCGAAACAAACAGCACAAATGCCATCGATGGTGCTACT-----	10263
Qy	278	AspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluValLys	297
		: :	
Db	10264	AACTTAAATAAGACACAAAAGAT-----CGGTAAACACCAAGTTACAAGTCGCCAA	10317
Qy	298	ArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyArgLeuGlyGlyLysIle	317
		: : : : : : :	
Db	10318	CGTGTGCAAAATGTAACAAGTATCCCAAACTGCA-----	10353
Qy	318	AlaTyAspTyTyAsnLeuPheAsnLysGlyTyIleGlySerValValTrpAspMet	337
		: : : :	
Db	10354	-----AATGAACTTAATACAGCT---ATGGGTCAATTACACATGGTATT	10395
Qy	338	AspLysTyThrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyArgGlyAsn	357
		: : : :	
Db	10396	GATGATGAAACGCCAAC-----AAACAACTCAAAAATATCGTCAGCCT	10440
Qy	358	TyrTrpThrSerAsnValSerTy-----Asn	366
		::: : : : :	
Db	10441	GAACAAGATGAACAACTGCTTATGATCAAGCTGTAGCTGCGAAGCAAAATTTAAAT	10500
Qy	367	ArgSerThrThrGlnAsnLeuGluTyArgAlaPheSerGlyGlyIleTyTyVal---	385
		: : : :	
Db	10501	AAACAACCTGGTTCAAATTCAGATAAGACAGCTTGACCGTGCATTTACACACAGTAACA	10560
Qy	386	-----ArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGly	403
		: : : : : : : :	
Db	10561	AGTACGAAGATGCATTCATGGCTATCGCGAACTA---CGAGAAGCGAAACGCCAGCT	10617

```

Qy 404 ArgYsIleProGlySerAspIleAsnSerHisAlaThrMetLeuThrAla 423
      ::::: |||::: ::::: |||::: |||::: |||::: |||::: |||:::
Db 10618 AAACAACACTTAGGTACTTTAAACACCATATTAAACATGCACACGCTACTCGTTAGCAAGGT 10677
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 424 SerTrpIysArgGlnLeu-----LeuAsnAsnValLeuHlsProGluAsnGly 439
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10678 CAATCAATCAACGCGACGCTGTGATGCGCTTAATACTGTG-----AAACAATGCC 10731
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 440 HisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10732 AATACATATTAGACGGCGCAATGAATAGCTTGCACAGGTTCAATCAATGATAAGATCGACA 10791
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 460 IleArgThrSerAlaArgAlaGlyTyrPhePheThrProGluAsnLysLys-----476
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10792 TTAAGAAATCAA-----AATATCTTCATCCAGATGATCAACAGTACCGCATAT 10842
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 477 -----LeuGlyThrPheIleIleArgGlyGlnAlaGlyTyrThrValAla 491
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10843 ACTCAAGCTGTCACAGCAGCTCAAGGCATTTTAAATAAACAACACTGGTGTAAACATCT 10902
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 492 ArgAspAsnAlaAspValProSerGlyLeuMetPheArgSer-----GlyGly 507
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10903 AAAGCAGAGCTTGATAATGCATTAAATACAGTTACAGACGGAACAGCGGCTTTAATGCT 10962
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySer 527
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 10963 GCTGAACACTTAAAGAAATACGAAACCTTCAGCAACAATAACGATT-----AATGGG--- 11013
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 528 ValLeuProGlu---ArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThr 546
      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 11014 ---TTACCAAACTTAACACATTTACAAAGATNACTTGAGCCTCAGTTGCACAAACGG 11070
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Qy 547 ArgThrLeuSerGly 551
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 11071 CAAATGTAGCAGGT 11085
      ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

RESULT 37
US-09-782-378A-26
; Sequence 26, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral vectors
; FILE REFERENCE: STONVB-04970
; CURRENT APPLICATION NUMBER: US/09/782, 378A
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 35100
; TYPE: DNA
; ORGANISM: Human adenovirus type 17
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (25184)..(25184)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
US-09-782-378A-26

```

Alignment Scores:	
Pred. No.:	10.2
Score:	100.00
Percent Similarity:	39.22%
Best Local Similarity:	25.10%
Query Match:	3.15%
DB:	10
Length:	35100
Matches:	64
Conservative:	36
Mismatches:	101
Indels:	54
Gaps:	14

US-09-857-669-2 (1-609) x US-09-782-378A-26 (1-35100)


```
Qy 369 ThrThrGlnAsnLeuGluLysArgAlaPheSerGly-----GlyIleTrpTyr 384
Db 1050 GCGAACCAAAATG-----CGGGATTTGAAGTTTATCGTTGAAGGCCCTCGAGTAT 1103
Qy 385 ValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyArg 404
Db 1104 TTAACGGTGTTCGAGC-----GAGACAGCAGCAAGGAATTACAACAATTCATTAT 1154
Qy 405 LysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSer 424
Db 1155 AAA---CCAGGGCGAAGACGAT-----GGCCTTATG 1184
Qy 425 TrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGlyHis-----440
Db 1185 TGGGGCGACTATCGATGAACCGGATT-----GAGCCAGGCAATCGTCTCGATCGC 1235
Qy 441 -----TyrLeuAspGlyLysIleGlyThrLeuGlyThr-----452
Db 1236 TTTTACAGGGGTTCGATCTGCGATGTGAAGCCCTCTGCGATTTTTCGACAGGC 1295
Qy 453 PheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrPro 472
Db 1296 TATTACACAAGAACCACTTGT-----GCTATGATTTGGACGGC 1337
Qy 473 GluAsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThr---ValAla 491
Db 1338 GAACAACCTGCACAGAAATGCTTTGTT-----GACAGCGGCTGGACCCGATGGAA 1388
Qy 492 ArgAspAsnAlaAspValProSerGlyLeu-----MetPhe 503
Db 1389 AACCCATTTTCACGATTTCACACATGCGGTAGACGGGAACCAATGACGAATAGCCACCTC 1448
Qy 504 ArgSerGlyGlyAlaSerValArgGlyTyrGluLeuAspSerIleGly-----520
Db 1449 ACCACAGGGGCACATTCATGACTGTTCCGATGTCGACGGGGATGGCAACATGAG 1508
Qy 521 ---LeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 1509 ATTTATTTATGTT-----GCTGCACGATTGACGATGACGGGACGCTCTGTATAGTCG 1562
Qy 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMetGlyAsp 559
Db 1563 TTTGCTGAATTCCT-----CCCGAA 1583
Qy 560 AlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyValArgTrp 579
Db 1584 AGTCAGCTCCAGGTGAATGGCAGGATTAGCCCATAGCCGATGCGCTTCATGTTGCCGAT 1643
Qy 580 PheSerPro-----LeuAla 584
Db 1644 ATTGATCCAGACAGCCAGGTTGGAATCATTTATGTTTGTGAAGGGGCCAGTATGG 1703
Qy 585 ProPheSerPhe-----AspIleAlaTyrGly 593
Db 1704 CCGTATGCTTTCGCTTTACGCGATGCGCGGTACAGCCGATGTCCTTTACGGA 1754

RESULT 39
US-09-938-842A-1039
; Sequence 1039, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
```

```
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1039
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1039

Alignment Scores:
Pred. No.: 0.274 Length: 3078
Score: 99.50 Matches: 113
Percent Similarity: 31.83% Conservative: 70
Best Local Similarity: 19.65% Mismatches: 225
Query Match: 3.14% Indels: 167
Db: 9 Gaps: 30

US-09-857-669-2 (1-609) x US-09-938-842A-1039 (1-3078)
Qy 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 121 TTTCACTCTACAATCTACTCAAAATCAAAGGCAGAGTCTGCT-----GCACCT 165
Qy 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 166 GTTCCACGTCT-----GTCCACATCTTCTAACTAACTAGATAGCTTCTTATAGAT 213
Qy 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySer 168
Db 214 CGAACAAAGCAGTCTGATCTAGAGGAGTTACAAGAGCTTGGGAGCTGATCCCAACAGT 273
Qy 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArg 188
Db 274 -----GTTGATGAGTCGCGGGATAACTTTTTAGGAATTTTGGCGTCAG-----318
Qy 189 LysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla 208
Db 319 -----GCTTCTACATGCGCTGGTATCTCGGGGCAACCACTT 354
Qy 209 ThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPheGlu 228
Db 355 CAAGAAAGCATGCGTTGTTG-----TTGCTAGTTAGAGCTTACCAG 396
Qy 229 IleThrGly-----ThrGlnArgTyrPro 236
Db 397 GTTAATGCCACACATGAAGCCCAAGCTTGATCTTATAGCTCTAGACAAGAGAGATTCCA 456
Qy 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 256
Db 457 GAGATCTCAGCCAGCTCTTTATGGGTT-----ACTGAGGCTGATCTTGATCGG 507
Qy 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
Db 508 GAATTC---TTTCTGGGTGTATGGAGGATGTCGGGTTTTTCTCTCT-----549
Qy 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
Db 550 -----GAGAACCCCGCGTT-----CAACA 570
Qy 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyLys 316
Db 571 CTGAGTCTGATCTGCTG-----AGGCTTGACCAAGCTTAC-----TGTGGGACT 615
Qy 317 IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
Db 616 ATAGGTTATGATGATCATGTCATCTGCTGATAGGATAAATGTAAC-----TGGTTG 666
Qy 337 MetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 667 AGACAGAAGATCGAGACC-----CCAACTCTCCGACAGTAC-----702
Qy 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
```



```
Db 703 AATAGTGGCGTGGATGTTATTATATAGG---CTTACCTGGAGCACACAGTTTGAG 759
QY 377 AlapheSerGlyIleTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 760 AATTTCTTGCTTACTAAGTGCACCGCTTAAAGAGTTTGACTGAA-----GGT 810
QY 397 AlaGluPheLeuAlaGluGlyArgLys-----IleProGly 408
Db 811 GCTGAATCTTTGATCTCGGATGAAGAGATGTTTCGATAGGTCTGCAGATCTCGGGTA 870
QY 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgIn 428
Db 871 GAGAACATAGTATCGGTATGCCCATAGGGTCGACTT-----AAT 912
QY 429 LeuLeuAsnValLeuHis----- 435
Db 913 GTTTGGGTAAATGTTGTAGAAACCTACGCCAAATATTACGCAGTTTACCGGTGT 972
QY 436 -----ProGluAsnGlyHisTyrLeu---AspGlyLysIleGlyThrThrLeu 450
Db 973 ACTAGGCCAGTAGATGAAGTTGGGCTTTACACCGGAACAGGTGATGTGAATACCACCTTG 1032
QY 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1033 GTATCATCTTATGATCGTCCAACTAGAGGAGGCAACATCTCCACTTGTCTTTGGTAGCA 1092
QY 471 ThrProGluAsnLysLysLeuGlyThrPheIleArgGlyGlnAla-----Gly 487
Db 1093 AATCCGAGTCACTTGGAGCAAGTAGATCCTGTTGTGATAGTAAATACCAGACGAAACAA 1152
QY 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
Db 1153 TATTACAGAAAGCAGCAACAGCAAGAAACATGGGTATTTTGATCCATGGGATGGT 1212
QY 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGly---ProAsn--- 525
Db 1213 AGCTTTGGCGGACAAAGGAGTGGTGTATGAACCTCTCCATCTTAGTGCACCTTCCCTAAC 1272
QY 526 -----GlySerValLeuProGluArgAlaLeuLeuValGly----- 537
Db 1273 TGTACCGGTGGNACAGTC-----CACATGTGTGTGAATCAATCAAGTGGCTTTC 1320
QY 538 -----SerLeuGlyTyrGlnLeuProPheThrArgThrLeu 549
Db 1321 ACAACCGATCCCGAGGAAGAGTCTTCACAGTATTTCACATGCTTTCGCAAGGCTTTC 1380
QY 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
Db 1381 ACGGCCCCAATTTCCATGTCATGCGAGATGACATTGAA-----GCAGTA 1425
QY 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
Db 1426 GTGCATGCTTGTGAGCTTGCCTCGCTAGTGG-----CGCCAGACGTTCCATTCGTAT 1476
QY 590 IleAla-----TyrGlyHisSerAsp----- 596
Db 1477 GTTGTCTTGATTAGTATGCTACCGCTTGGGGATGAACGAGATACACGAACCGTCA 1536
QY 597 -----LysLysIleArgTrpHisIleSer 604
Db 1537 TTCACACACCAAAAAATGTACAAGGTGATACGCGAGTCACTCCCTCG 1581
```

RESULT 40

```
US-09-897-056-6
; Sequence 6, Application US/09897056
; Patent No. US20020100074A1
; GENERAL INFORMATION:
; APPLICANT: JOE, YUJI
; APPLICANT: MIMA, TETSUYA
; APPLICANT: KIDA, TAKAO
; APPLICANT: IGARASHI, DAISUKE
; APPLICANT: OHSUMI, CHIEKO
```

```
; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSFORMED PLANT HAVING INCREASED GLUTA
; FILE REFERENCE: 210425USO
; CURRENT APPLICATION NUMBER: US/09/897,056
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: JP 2000-215279
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 3412
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(3152)
; OTHER INFORMATION:
US-09-897-056-6

Alignment Scores:
Pred. No.: 0.321 Length: 3412
Score: 99.50 Matches: 113
Percent Similarity: 31.83% Conservative: 70
Best Local Similarity: 19.65% Mismatches: 225
Query Match: 3.14% Indels: 167
DB: 10 Gaps: 30
```

US-09-857-669-2 (1-609) x US-09-897-056-6 (1-3412)

```
QY 109 PheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHisIleThrPro 128
Db 198 TTTCACTCTACAACTACTCAAAATCAAGGCAGAGTCGTCT-----GCACCT 242
QY 129 GlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAsp 148
Db 243 GTTCCACGCTCCT-----GTCCCACTTTCTAAGCTAACTGATAGTCTTCTTAGAT 290
QY 149 GlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnProValGlySer 168
Db 291 GGAACACGACGTGTCTATCTAGAGGAGTTACAAAGAGCTTGGGAGGCTGATCCCAACAGT 350
QY 169 AspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArg 188
Db 351 -----GTTGATGAGTCTGGGATAACTTTTTTAGGAATTTTGTGGGTGAG----- 395
QY 189 LysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAla 208
Db 396 -----GCTTCTACATCGCTGCTATCTCGGGCAAAACCAT 431
QY 209 ThrValAspLeuAsnValValValAspSerGlyArgProIleAlaPheGlyAspPheGlu 228
Db 432 CAAGAAGCATCGCTTTGTTG-----TTGCTAGTTAGAGCTTACCAG 473
QY 229 IleThrGly-----ThrGlnArgTyrPro 236
Db 474 GTTAATGGCCACATCAAGCCCAAGCTTGATCCTTTAGTCTACAGAAGAGAGAGATTCCA 533
QY 237 GluGlnIleValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeu 256
Db 534 GAGGATCTCAACCCCGAGGTCTTTATGGGTTT-----ACTGAGGCTGATCTTGTACGG 584
QY 257 LeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGln 276
Db 585 GAATTC---TTTCTGGGTGATGGAGGATGTCGGGTTTCTCTCT----- 626
QY 277 AlaAspPheAspArgLeuGlnGlyAspArgValProValLysValSerValThrGluVal 296
Db 627 -----GAGAACCGCCCGGTT-----CAACA 647
QY 297 LysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLys 316
Db 648 CTCAGGTCGACTACTCTCG-----AGGCTTGAGCAAGCTTAC-----TGTGGGACT 692
```

Search completed: November 9, 2002, 03:59:22
Job time : 497 secs

```
Qy 317 IleAlaTyrAspTyrTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAsp 336
Db 693 ATAGGCTATGAGTACATGCACATTCGTGATAGGATAAATGTAAC-----TGCTTG 743
Qy 337 MetAspLysTyrGluThrThrLeuAlaGlyIleSerGlnProArgAsnTyrArgGly 356
Db 744 AGACACAAGATCGAGACC-----CCAACTCCTCGACAGTAC-----779
Qy 357 AsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArg 376
Db 780 AATAGTGCAGCGTCGGATGGTTATTTATGATAGG---CTTACCTGGAGCACACAGTTTGAG 836
Qy 377 AlaPheSerGlyGlyIleThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGly 396
Db 837 AATTCTTGCTACTAAGTGGACCGGCTAAAGGTTTGGACTGGAA-----GGT 887
Qy 397 AlaGluPheLeuAlaGluGlyArgLys-----IleProGly 408
Db 888 GCTGAATCTTTGATTCCTGGCATGAAGAGAGATGTTTCGATAGGTCTCGCAGATCTCGGGGA 947
Qy 409 SerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSerTyrLysArgGln 428
Db 948 GAGACATAGTTATCGGTATGCCCATAGGGTCGACTT-----AAT 989
Qy 429 LeuLeuAsnValLeuHis-----435
Db 990 GTTTTGGGTAAATGTTCTTAGAAAACCTCTACGCCAAATATTCAGCGAGTTTAGCGGTGGT 1049
Qy 436 -----ProGluAsnGlyHisTyrLeu---AspGlyLysIleGlyThrThrLeu 450
Db 1050 ACTAGGCCAGTAGATGAGTTGGGCTTTACACCGGACAGGTGATGGAATACCACCTTG 1109
Qy 451 GlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePhe 470
Db 1110 GGTACATCTTATGATCGTCCAACTAGAGGAGGCAAAACATCTCCACTTGTCTTTGGTAGCA 1169
Qy 471 ThrProGluAsnLysLysLeuGlyThrPheIleIleArgGlyGlnAla-----Gly 487
Db 1170 AATCCAGTCACCTTGGACGACTAGATCCTGTTGTGTAGTAAACCCAGACGAGAAACAA 1229
Qy 488 TyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGly 507
Db 1230 TATTACACGAAAGACGAGACAGAACAAAGAACATGGGTATTTGTATCCATCGGGATGGT 1289
Qy 508 AlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGly---ProAsn--- 525
Db 1290 AGCTTTGCCGACAGGAGTGTCTATGAAACTCTCCATCTTACTGCACCTTCCCTAACTAC 1349
Qy 526 -----GlySerValLeuProGluArgAlaLeuLeuValGly-----537
Db 1350 TGTACCGGTGGAACAGTC-----CACATTGTGTGTAATCAAGTGGCTTTC 1397
Qy 538 -----SerLeuGluTyrGlnLeuProPheThrArgThrLeu 549
Db 1398 ACAACCGATCCCGAGGAGGAGGCTTTCACAGTATTGCACGTATGCTTGCAGAGGCTTTC 1457
Qy 550 SerGlyAlaValPheHisAspMetGlyAspAlaAlaAsnPheLysArgMetLysLeu 569
Db 1458 AGCGCCCCCAATTTTCCATGTCATGACATGACATTTGAA-----GCAGTA 1502
Qy 570 LysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPheAsp 589
Db 1503 GTGCATGCTTGTGAGCTTGCCTCAGTGG-----CGCCAGACGTTCCATCTCTGAT 1553
Qy 590 IleAla-----TyrGlyHisSerAsp-----596
Db 1554 GTTGTGTTGTTGTTATGATGTCGCTTGGCGATTAACGAGATAGACGAAACCGTCA 1613
Qy 597 -----LysLysIleArgTrpHisIleSer 604
Db 1614 TTCACACAAACCAAAATGTACAAAGGTATACGCGAGTCATCCCTCG 1658
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:14:59 ; Search time 32 Seconds
(without alignments)
3921.334 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPPHAY.....IAGHSDKKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_todent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriaph:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3153	99.4	615	Q9JWM5	Q9JWM5 neisseria m
2	3136	98.8	615	Q9JXB7	Q9JXB7 neisseria m
3	596	18.8	589	Q8XWT6	Q8XWT6 raietonia s
4	449	14.2	579	Q910U1	Q910U1 pseudomonas
5	368.5	11.6	578	Q8ZB95	Q8ZB95 yersinia pe
6	355.5	11.2	577	Q8XEL5	Q8XEL5 salmonella
7	348	11.0	582	Q9KFP3	Q9KFP3 vibrio chol
8	340	10.7	618	Q92LN9	Q92LN9 rhizobium m
9	336	10.6	586	Q9CK26	Q9CK26 pasteurella
10	332.5	10.5	765	Q8X2I3	Q8X2I3 raietonia s
11	323.5	10.2	641	Q8UC83	Q8UC83 agrobacteri
12	306.5	9.7	617	Q9PD27	Q9PD27 xyella fas
13	300	9.5	738	Q9X3V8	Q9X3V8 zymomonas m
14	292	9.2	593	Q9EY32	Q9EY32 xanthomonas
15	290.5	9.2	623	Q8YEI3	Q8YEI3 brucella me
16	279	8.8	833	Q8YUR6	Q8YUR6 anabaena sp

17	272.5	8.6	628	16	Q9A7W6	Q9A7W6 caulobacter
18	260.5	8.2	792	2	P95359	P95359 neisseria g
19	253	8.0	797	16	Q9KIH0	Q9KIH0 neisseria m
20	253	8.0	797	16	Q9JX31	Q9JX31 neisseria m
21	251	7.9	797	2	O30912	O30912 neisseria m
22	246	7.8	778	16	O67326	O67326 aquifex aeo
23	246	7.8	795	16	O8ZH58	O8ZH58 yersinia pe
24	240.5	7.6	617	16	Q98K30	Q98K30 rhizobium l
25	228.5	7.2	804	16	O8ZRP0	O8ZRP0 salmonella
26	227	7.2	797	2	O9S341	O9S341 photorhabdu
27	227	7.2	803	16	Q9KFW0	Q9KFW0 vibrio chol
28	226	7.1	803	16	O8Z9A3	O8Z9A3 salmonella
29	219	6.9	792	2	O32625	O32625 haemophilus
30	217	6.8	785	2	O9R2E3	O9R2E3 escherichia
31	215.5	6.8	797	16	O9HXY4	O9HXY4 pseudomonas
32	211	6.6	676	16	O8YWP0	O8YWP0 anabaena sp
33	210.5	6.6	769	16	O9A711	O9A711 caulobacter
34	206	6.5	739	16	O9P1Z8	O9P1Z8 campylobact
35	202.5	6.4	861	16	P73472	P73472 synecocyst
36	199	6.3	792	16	O9PKF3	O9PKF3 chlamydia m
37	197.5	6.2	784	16	O9PEI2	O9PEI2 xyella fas
38	194.5	6.1	781	16	O8YHH0	O8YHH0 brucella me
39	193.5	6.1	793	2	O93PM2	O93PM2 haemophilus
40	189.5	6.0	776	16	O92O48	O92O48 rhizobium m
41	185	5.8	774	16	O8UFL6	O8UFL6 agrobacteri
42	179	5.6	491	16	O8ZOL6	O8ZOL6 anabaena sp
43	179	5.6	789	2	Q51930	Q51930 pasteurella
44	178.5	5.6	794	16	Q98MC3	Q98MC3 rhizobium l
45	177	5.6	678	16	O8RHT4	O8RHT4 fusbacteri

ALIGNMENTS

RESULT 1

Q9JWM5	ID	Q9JWM5	PRELIMINARY;	PRT;	615 AA.
AC	Q9JWM5;				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DPC-2001 (TREMBLrel. 19, Last annotation update)				
DE	Putative outer membrane protein.				
GN	NNA0296;				
OS	Neisseria meningitidis (serogroup A).				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=65699;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;				
RX	MEDLINE=20222556; PubMed=10761919;				
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,				
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,				
RA	Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,				
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,				
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,				
RA	Whitehead S., Spratt B.G., Barrell B.G.;				
RT	*Complete DNA sequence of a serogroup A strain of Neisseria				
RT	meningitidis 22491*;				
RL	Nature 404:502-506(2000).				
DR	EMBL; AL162752; CAB83602.1; -.				
KW	Complete proteome.				
SO	SEQUENCE 615 AA; 67775 MW; B9A6460763BE1CCB CRC64;				

Query Match	99.4%;	Score 3153;	DB 16;	Length 615;
Best Local Similarity	99.3%;	Pred. No. 2.7e-192;		
Matches	605;	Conservative	1;	Mismatches 3;
			Indels	0;
			Gaps	0;
Oy	1	MMKPTALLPALFFPPHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ	60	
Db	7	MMKPTALLPALFFPPHAYAPADLSENKAAGFALFNKSPDTSVKLKPFPVRIDTQ	66	
Oy	61	DSEIKDWEHLPLITQOOEEVLDEKGTGFLAEAPDNVKTMLRSKGVFSSKVSLSLTGKDG	120	

```

Db 67 DSEIKDMVEEHLPLITQQOEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLSLEKDG 126
QY 121 AYTWHITPGPRTKIANGVVAIIIGDILSDGNLAEEYRNALENNWOQPVGSDFDQDSWENSKT 180
Db 127 AYTWHITPGPRTKIANGVVAIIIGDILSDGNLAEEYRNALENNWOQPVGSDFDQDSWENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDSDGRPIAFGDFEITGTQRYPQIV 240
Db 187 SVLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVVDSDGRPIAFGDFEITGTQRYPQIV 246
QY 241 SGLARFOPGTPYDLDLLDFQALBONGHYSCASVQADFDRLQGDVRPVKVSVEYKRHK 300
Db 247 SGLARFOPGTPYDLDLLDFQALBONGHYSCASVQADFDRLQGDVRPVKVSVEYKRHK 306
QY 301 LETGIRLSDSEYGLGGKIATDYNNLKNKGYIGSVVWMDKYEYTTLLAAGISQPRNRYGNWT 360
Db 307 LETGIRLSDSEYGLGGKIATDYNNLKNKGYIGSVVWMDKYEYTTLLAAGISQPRNRYGNWT 366
QY 361 SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 426
QY 421 LTASWKRLNLYLHPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKGTF 480
Db 427 LTASWKRLNLYLHPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKGTF 486
QY 481 IIRGOAGYTVARNADVPGLMFRSGGASSVRYELDSIGLAGPNSVLPERALLVGSLE 540
Db 487 IIRGOAGYTVARNADVPGLMFRSGGASSVRYELDSIGLAGPNSVLPERALLVGSLE 546
QY 541 YQLPFTRTLSCGAYFHDMDGAAANFRMKLKHGSLGVRWFSPLAPFSDIAYGHSKKIR 600
Db 547 YQLPFTRTLSCGAYFHDMDGAAANFRMKLKHGSLGVRWFSPLAPFSDIAYGHSKKIR 606
QY 601 WHISLGTFRF 609
Db 607 WHISLGTFRF 615

RESULT 2
Q9JXB7 PRELIMINARY; PRT; 635 AA.
ID Q9JXB7
AC Q9JXB7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Hypothetical protein NMB2134.
GN NMB2134.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Kaulchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Winn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Colton M.D., Urtabek T.R., Khouri H., Qin H., Yamathevan J.,
RA Gill J., Scarlat V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002561; AAF42442.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 635 AA; 69801 MW; 2B971F47A9DDA98 CRC64;

Query Match 98.8%; Score 3136; DB 16; Length 635;
Best Local Similarity 98.7%; Pred. No. 3.4e-191;

```

```

Matches 601; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 MMKPTALLPALFPFPFHAYAPAAOLSENKAAGFALPKNKSPTDESVKLPKFPVRLDQ 60
Db 27 MMKPTALLPALFPFPFHAYAPAAOLSENKAAGFALPKNKSPTDESVKLPKFPVRLDQ 86
QY 61 DSEIKDMVEEHLPLITQQOEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLSLEKDG 120
Db 87 DSEIKDMVEEHLPLITQQOEVLDEKQTFGLAEAPDNVKTMLRSKGYFSSKVSLSLEKDG 146
QY 121 AYTWHITPGPRTKIANGVVAIIIGDILSDGNLAEEYRNALENNWOQPVGSDFDQDSWENSKT 180
Db 147 AYTWHITPGPRTKIANGVVAIIIGDILSDGNLAEEYRNALENNWOQPVGSDFDQDSWENSKT 206
QY 181 SVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDSDGRPIAFGDFEITGTQRYPQIV 240
Db 207 SVLGAVTRKAYPLAKLGNTRAAVNPDTATADLNVVVDSDGRPIAFGDFEITGTQRYPQIV 266
QY 241 SGLARFOPGTPYDLDLLDFQALBONGHYSCASVQADFDRLQGDVRPVKVSVEYKRHK 300
Db 267 SGLARFOPGTPYDLDLLDFQALBONGHYSCASVQADFDRLQGDVRPVKVSVEYKRHK 326
QY 301 LETGIRLSDSEYGLGGKIATDYNNLKNKGYIGSVVWMDKYEYTTLLAAGISQPRNRYGNWT 360
Db 327 LETGIRLSDSEYGLGGKIATDYNNLKNKGYIGSVVWMDKYEYTTLLAAGISQPRNRYGNWT 386
QY 361 SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 387 SNVSYNRSTTQNLKRAFSGGIWYVDRAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 446
QY 421 LTASWKRLNLYLHPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKGTF 480
Db 447 LTASWKRLNLYLHPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFTPENKKGTF 506
QY 481 IIRGOAGYTVARNADVPGLMFRSGGASSVRYELDSIGLAGPNSVLPERALLVGSLE 540
Db 507 IIRGOAGYTVARNADVPGLMFRSGGASSVRYELDSIGLAGPNSVLPERALLVGSLE 566
QY 541 YQLPFTRTLSCGAYFHDMDGAAANFRMKLKHGSLGVRWFSPLAPFSDIAYGHSKKIR 600
Db 567 YQLPFTRTLSCGAYFHDMDGAAANFRMKLKHGSLGVRWFSPLAPFSDIAYGHSKKIR 626
QY 601 WHISLGTFRF 609
Db 627 WHISLGTFRF 635

RESULT 3
Q8XWT6 PRELIMINARY; PRT; 589 AA.
ID Q8XWT6
AC Q8XWT6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Putative lipoprotein.
GN RSC2384 OR RS02766.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlegel T.,
RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissensbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646069; CAD16091.1; -.

```

[illegible]


```

Db 337 TDLNDTEODSTTLAVSRWDLSSGNORAI---NLRWSEDFHTQGNVTNTMTLFFPGVMIS 393
Qy 461 RTSARAG-----YFFPENKKLGT-----FIIRGOAGY 488
Db 394 RTRSRGLPTMGDSQRYSDVSNTPAGSDYDFSLQAOQNVWITFLVDRHFRVNRANLWG 453
Qy 489 TVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNCISVLPERALLVGSLEYQLPFTT 548
Db 454 IETGDFDKVPPDLRFAGGDSIRGYKYSIPKSDSDGNLKGASKLATGSLGYQYNVTGK 513
Qy 549 LSGAVFHMGDAAANFKRMKLKHGSGLVRFWSPLAFSFDIAGHSDKK---IRWHISL 605
Db 514 WKGAVFVDSGEAVSDIRSDFTGTGTGVGRWASPVGVKLDFAVPVGVGDKDEHGLQFYIGL 573
Qy 606 G 606
Db 574 G 574

RESULT 7
Q9KP31
ID Q9KP31 PRELIMINARY; PRT: 582 AA.
AC Q9KP31;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Hypothetical protein VC2548.
GN VC2548.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406633; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Mayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004323; AAF95689.1; -
DR TIGR; VC2548; -
DR InterPro; IPR000184; Bac_surfAg_D15.
DR InterPro; IPR000566; Lipocin_cytrABP.
DR Pfam; PF01103; Bac_surface_Ag_1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 582 AA; 65644 MW; E96946F8B5F1C426 CRC64;

Query Match 11.0%; Score 348; DB 16; Length 582;
Best Local Similarity 23.1%; Pred. No. 6.8e-14;
Matches 136; Conservative 96; Mismatches 255; Indels 102; Gaps 19;

Qy 65 KDWVEEHLPLITQO-----EEVLDEQGTGFLAEEAPDNVTKMLRSKGYFSKVSLT- 116
Db 46 KDWVEEHLPLITQO-----EEVLDEQGTGFLAEEAPDNVTKMLRSKGYFSKVSLT 93
Qy 117 -EKDGVTHVTPGTRTKIANGVAILGDILSDGNLAERYNALENKQOQVPGSDFDQSDS 175
Db 94 SEDNQRLRAAVTLGVTLSLSEVDIVIRGEAGDRDQRLIRRS---GLRVDAPLNHSLY 149
Qy 176 ENSKTSVLGAVTRKGYPLAKLGNTRAAVNPOTATVDLVNVDVDSGRPTAFGDFEITGORY 235
Db 150 DNLKSGIRNLALQKFGNGDFOASLEVIPELNQARVILHFDSGIRYLFATTVEGSO-I 208
Qy 236 PEQIVSGLARFQGTPTDLDLLDLDLQQAELQNGHYSGASVQADFRL-QGDRVPKVSVT 294
Db 209 DENRVMSLRPFKQGEYLYSQVNGEENQNLNTDWFSSVVEFVPELDLSQLDEGLRPIKTLA 268

```

```

Qy 295 EVRRHKLETGIRLDSYGLGGKIAYDYNNLKNKYIGSVVMDMKYETTLAAGISOP--- 351
Db 269 POARNOLETGLGYSTDGVGRSLKWKPKWNSQHSFSDSLSPBQTTITAGYKIPLED 328
Qy 352 ---RNTYRGNVTSNVSYNRSTONLEKRAFSGGIWVRDRAGIDARLGAEBFLAEGRKIPG 408
Db 329 ALNEYVRIQYGMKHL--DKRDTESLESNLSLRRHW----- 361
Qy 409 SDIDLNSHATMTATSKKROLLNNV---LHPENGHY-LDGKIGTTLGT----- 452
Db 362 -OLD-GGWHTVFI-----RYLLENYROGLQDDNSQFLPLGMYTTRTTRTSNSGLLTWGDK 415
Qy 453 -----FLSSALIRTSARAGYFTT-PENKKLTGTIIIRGOAGYTVARNADVP 500
Db 416 QTITLEYGDPAALLSETVRLQTSGLTSSWLTARNHRA---LVRVDGGANLVDFDQLSPS 472
Qy 501 LMFPSGASVVRGYELDSIGLAGPNCISVLPERALLVGSLEYQLPFTTSLGAVFHDGMDA 560
Db 473 LRFFAGGDNLRGYGYSIPKSDSDGNLKGASKLATGSLGYQYNVTGK 532
Qy 561 AANFKRMKLKHGSGLVRFWSPLAFSFDIAG---HSDKKIRWHISL 606
Db 533 FND--NPEKKKGVTGIRWISPVGPIRLDFAWGLDAAPGDEFKIHFTLG 579

RESULT 8
Q92LN9
ID Q92LN9 PRELIMINARY; PRT: 618 AA.
AC Q92LN9;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Hypothetical protein R02996.
GN R02996 OR SMC03097.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=213996507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591792; CAC47575.1; -
DR InterPro; IPR000184; Bac_surfAg_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; Bac_surface_Ag_1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 618 AA; 65656 MW; E51276BB29B1FB78 CRC64;

Query Match 10.7%; Score 340; DB 16; Length 618;
Best Local Similarity 22.0%; Pred. No. 2.4e-13;
Matches 144; Conservative 102; Mismatches 266; Indels 144; Gaps 22;

Qy 19 AYAPADLSENKAAGFALPKNKSPTDESVKLAKPEPVRI-----DFQDSIKDMVEHL 72
Db 2 ASAVLGPVSVEOAHAFKIFGRFFESAEVEQVIDPVRYTLTTFEFGTDEELREALNGS 61
Qy 73 PLITQOQEEVLDKEQGTGFLAEEAPDNVTKMLRSKGYFSKVSLSL-----TEKQAY- 122
Db 62 OLVDQOEEVSGDGLIAGIKARDDRRLAVLYEKARYGTVSILVNGODIDSLPPDPAFP 121

```

```
QY 123 -----TVHTPGPRTKIANGVAILGDILSDGNLAERYNALENMPOQVGSDFDQD 175
DB 122 DGQPVVVRVAFCP-----AFTLGTVRLEGDAARL-----DPAAYDLKRGAR 164
QY 176 ENSK-----TSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDVSGRPIAFGDFEI 229
DB 165 ADSTLIKAGEQIVNDLKEQSRPLAKLAERSVVAADHATSTVDVTIRADGGVAPVGNLT 224
QY 230 TGTORYPEQIVGLARFPQCTPYDLDLLDFOALEQNGHYSGASVQADFRLQDGRVPKYSV 288
DB 225 SGARTVDPDFVKDYSRLNHGRPSNPINRKAERLQNLVFSSTIN-EADGLAPDGTIP 293
QY 289 VKSVTEVKRHKLETGIRLDSYGLGGKIAYDYNNLFNKGYIGSVVMDMDKY-ETTLAAG 347
DB 284 MNIQVSECKHRYFGFGQVSTTQGLGLOGYWGHRNLFGRAESLRLEGSDRLGETTQVAG 343
QY 348 ISQPRNRYGNWTSNVSYNRSTQNLKRAFGSGGIWYVRDRA----- 389
DB 344 LD-----YSAGILFAKPGAFGPASTFTASVKAALVDP 375
QY 390 -----GIDARLGAEF-----LAERKIPGSDID-----LGNSHAT-MLTASWKROLLNN 432
DB 376 DAYSAKVTAAAGAAFELESPEDFSYGAEGVWADVDADFAGNSYITAAALPPEYVRDARD 435
QY 433 VLRHPENG-----HYLDGKIGTTLGTFTLSSTALIRT SARAGYFFTPENKKLGTFII 482
DB 436 KLNTEGYRALINAKPSYELEK-----TFESS-----FEASAGYAFGEKTR-----FVL 482
QY 483 RCQ--AGYTVARDN-ADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGSL 539
DB 483 ACKLAGVLYGGDELSDIPATRRFFLLGGGSGVRGYSYQISPRADDELITGGRSYVSGSL 542
QY 540 EYQLPFTRTLGSAGVFIDMG-----DAAANFKRMKLKHGSLGVNRFSP LAPPSFDIA 591
DB 543 EARIATVDTIGVVPFIDAGIVSDSTAPDS--DIRAGAGIGLRVATPFGPIRLDFA 596
RESULT 9
Q9CK26 PRELIMINARY; PRT; 586 AA.
ID Q9CK26;
AC Q9CK26;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Hypothetical protein PM1809.
GN PM1809.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70.
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006218; AAK03893.1;
DR InterPro; IPR00184; Bac_surfAg_D15.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF01103; Bac_surface_Ag; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 586 AA; 67106 MW; 00895174DED56283 CRC64;
Query Match 10.6%; Score 336; DB 16; Length 586;
Best Local Similarity 25.0%; Pred. No. 4e-13;
Matches 142; Conservative 88; Mismatches 292; Indels 46; Gaps 15;
QY 50 QDSEIKDWZEHPLPTIQOGEVLDKEQTGFLAEAPQNVKMTMLRSKGYFSKVSLSL----- 115
DB 41 QNTQLNNRYVGMII--DKEADGSGRHKQLVREADIKA---LRAGYYSQSEVEFOIES 95
```

```
QY 116 --TEKGAYTVHTPGPRTKIANGVAILGDILSDGNLAERYNALENMPOQVGSDFDQD 173
DB 96 QKPKKQOLIAHWKIGKPKYKIDVDVFOGEATQDFEIAVTKNIPEK-----GSIUNHIE 150
QY 174 SWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDVSGRPIAFGDFEITGTQ 233
DB 151 TYDNKSHIQIALSRGYFDGDFLVRLEVRPSTQOAWNRDLYDSGERYRFEVTFENAQ 210
QY 234 RYPEQIVGLARFPQCTPYDLDLLDFOALEQNGHYSGASVQADFRLQDGRVPKYSV 293
DB 211 -IREDYLRNNMINFKGOPYLINDLSTLTNNYTSSNWFSSVLMQPVLD-EHKIYNVDVLL 268
QY 294 TEVRKHKLETGIRLDSYGLGGKIAYDYNNLFNKGYIGSVVMDMDKYETTLAAG-----I 348
DB 269 QPRKKNSEVIGWASDVGRLOLWTKPWNNRHSRPTNLVYSAPKOTLEATYKMLPL 328
QY 349 SOPNRYGNWTSNVSYNRSTQNLKRAFGSGGIWYVRDRAIDARLGAELAEGRKIPG 408
DB 329 KNPINYYEVSAGLENNKNDTSPASSL--SAIRYNHNEAGWQHSIGLRVYD----- 380
QY 409 SDIDLGNHATML---TASWKROLLNVLPENGHYLDGKIGTTLGT--FLSSTALIRTS 463
DB 381 SFIOANVKDKTLLVFTASVRRTRLOGLFPTWGD--TQKLTIDLGRTWMLSDVDFLKM 438
QY 464 ARAGYFT-PENKKLGTFIIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLA 522
DB 439 GSSLVWRTYIQLHRI---VTRLEGLWHLTKNIERTPPALRPFAGGDSIRGYGKKIAPK 495
QY 523 GPNSGVLPERALLVGSLVYOLPPTRTLGSAGVFDHMGDAAAANFKRMKLKHGSLGVNRFSP 582
DB 496 NNAGKLVGGSRLITGSEFYQYQYVPDWMLATFADTGLAANOFTTKELRYGAGMGVRWASP 555
QY 583 LAPPSFDIAVGHSD---KKIRWHISLG 606
DB 556 VCAIKFDIATPVRDKNSKNIQIFYIGLG 583
RESULT 10
Q8XZ13 PRELIMINARY; PRT; 765 AA.
ID Q8XZ13;
AC Q8XZ13;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative outer membrane signal peptide protein.
GN RSC1412 OR RS05280.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000.
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Manganot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Welssnabach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646064; CAD15114.1; -.
DR InterPro; IPR002198; ADH_SHORT.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 765 AA; 85878 MW; 378641PB9P247C68 CRC64;
Query Match 10.5%; Score 332.5; DB 16; Length 765;
Best Local Similarity 24.5%; Pred. No. 1e-12;
Matches 140; Conservative 89; Mismatches 245; Indels 97; Gaps 23;
```



```
Qy 99 VKTMRKSGYF-----SSKVSUT--EKDGAVTVHTGPRRTKIANVGVAILGDILSDGML 151
Db 232 LRSFYLDGYLEFALESTOYSTDPDKDKYTLTLNHEGEQKYSD--IKLTGCELLS--KO 287
Qy 152 AEYRNALNNWQOPVGSDFDSDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPOTATVD 211
Db 288 AE-MEXLIKQDVFSSAKLSATTATKATDILGTY---GYAFAT-NPQPIQNKDRTVA 342
Qy 212 LNVVDSGRPIAFGDFEITGCTORYPQIVSGLARFQGTPTDLDLLDFOALQNGHYS 271
Db 343 LTVLVGRRVYVRRVNVVNSKTRDEYVRRMROMESSWFDGEKQLQSONRINRTGYFT 402
Qy 272 GASVQADFRLQCDRVPKVSYTE-----VRRHKLETGIRLDSYGLGGRI 317
Db 403 DTNITTEDVPGASDQVDVNVNVTERTQGISLGVGFSTDKLVLSAGIRODNVFGSSTSL 462
Qy 318 AYDYNLFNKGIGSVVWDMXYETTLAAGISQP-----RNYRGNWYTSNVSNRTTQN 372
Db 463 GLDVNTSKSNRTIS--VTOYDPYFT--VDGISRSTELYRYRPLYTYTGDDY-RIVQOG 517
Qy 373 -----LEKRAFSGGIWVRDRAGIDARLGA-----BFLAEGRKIPGSDIDLGNSHA 418
Db 518 GNIKFGVPSEDTDTVPFGIGY--ERTTIDVTSNPLVYQNVAKNGRI-----SNN 566
Qy 419 TMLTASKKROLLNNVLPENHGYLDGI-----GTTLTGTFSLTALRTSARAGY---FF 470
Db 567 FPIITGMSKDQDSDALSALVPTGRYQOANLEFGIPGGLQY-----RAYYQHGYF 615
Qy 471 TPNKKLGTFTIIRGOAGYTVARNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLP 530
Db 616 YPVSKSF-TLAFNEIGYGHGYNKDFPVFKNYVAGGIGSVRGYETSTLPDANGVAIG 674
Qy 531 ERALLVGSLEYQLPF-----TFTLSGAVPHMGDAAN---FKRMKLKHSGSLGVWRFSP 582
Db 675 GASKFVGNMEFTPLPGSGVDRTVRLFTFFDYGNVFARGQPKVGDMDRYSGAGFLSWLSP 734
Qy 583 LAP-----FSFDIAYGSHDKIRWHISLGTRF 609
Db 735 IGPLKISMGPPIKRAEDQTORFOIGTAF 765

RESULT 11
Q8UC83 ID Q8UC83 PRELIMINARY; PRT; 641 AA.
AC Q8UC83;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Hypothetical protein Atu2615.
GN ATU2615 OR AGR_C_4742.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavain T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.*;
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorllo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.*;
RL Science 294:2323-2328(2001).
DR EMBL; AB009208; AAL43596.1; ALT_INIT.
DR EMBL; AF001175; AAK88337.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 641 AA; 68281 MW; EC0A11DA3BBD61E CRC64;

Query Match 10.2%; Score 323.5; DB 16; Length 641;
Best Local Similarity 22.6%; Pred No. 2.9e-12;
Matches 148; Conservative 94; Mismatches 290; Indels 123; Gaps 21;

Qy 6 TALLLPALFFPHAYAPADLSENKAAGFALFKNKSPDTSVKLPKPPVRIPTQDSE-- 63
Db 19 TALAVALTEALYPAF--ARDAFAFKLFGMLWGSPEPEVEVIN-PVKYAVTLDAAADAKS 75
Qy 64 IKDMVEHLPLITQOOEEVLDEKQTFGL-----AEEAPDNVKTMLRSKGVFSSKVSJLT-- 116
Db 76 LKSSLENSLLIADK-----DKPASGDLGLIKARDDRDLIAALIENARYGGIVNVTVA 130
Qy 117 -----EKDGAVTVHTGPRRTKIANVGVAILGDILSDGSLNLAERYRNALEN 161
Db 131 GKNVDDLPPNPVFDHSTPVPVMTITVPGPKPTLGN--VRLEGDV-TGRNLDGLIA--- 184
Qy 162 WQOPVGSDFDQDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPOTATVDLNVVDSGRP 221
Db 185 -----GGDAGSLAIIRAGNKLIDDLKAEGRPLAKLTRKREAVANHATTVITWAAEGGPV 239
Qy 222 IAFGDFEITGCTORYPQIVSGLARFQGTPTDLDLLDFOALQNGHYSAGSVQADFDR 281
Db 240 APLGAVTVTGEKTVGDGFIRRYSLRNGEPYSPEKRAADRRLQGLGVFSSLTKEAGTL 299
Qy 282 LOGDRVPKVSYTEVKKHKLETGIRLDSYGLGGKIADYDYNLFNKG-----YIGSV--- 333
Db 300 ARDGTPLPIEVSSEGHRYFGVGAQYSTEIGIGLOGYGHENLFGQAESLRIGSVSRIA 359
Qy 334 -----VWMD-----KYETTLAAGISOPRNYRGNWYTSNVSNRTTQN 372
Db 360 EASSVEGMDYSAGITFTKPFMFPNPTTFTKSLIAKTENPDYTRAKTLTG----- 409
Qy 373 LEXRAFSGGIWVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASMKROLLNN 432
Db 410 -----AGFAYELND-----TDTAAGLEVQWADTEDAFCKNEYLTTSIPLEVRD 454
Qy 433 VLHPENHYLDGKIGTGLTGLTFLSSTALRTSARAGYFTTPEN-----KKLGT---PIIR 483
Db 455 T-----RDDKLNPTGFRASLAAPSPYEALNGTFFSFSPGSIYKGLGAEEDRLIMA 506
Qy 484 GQ--AGYTV-ARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLE 540
Db 507 GKLSGGVILVCGSDLDIPTTRRFAGGGGSGVSGYQELSPYNAAGDATGGRSVVGSVE 566
Qy 541 YQLPFTFTLSGAVPHMG-----DAANFKRMKLKHSGSLGVWRFSPPLAPSFDLA 591
Db 567 ARIKVTDTIGLVFFFDAGVYVSDGVTPDFS--DIRAGAGIGLRYATPTFGPLRLDVA 619

RESULT 12
Q9PDZ7 ID Q9PDZ7 PRELIMINARY; PRT; 617 AA.
AC Q9PDZ7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein xf1231.
GN xf1231.
```

```

OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reznach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber J.P.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lenos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.N.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.E.K., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000)
DR EMBL: AE003957; AAF84041.1; -
DR InterPro: IPR000184; Bac_surfaG_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 617 AA; 69579 MW; A875FA462C470D21 CRC64;

Query Match 9.7%; Score 306.5; DB 16; Length 617;
Best Local Similarity 23.3%; Pred. No. 3.3e-11;
Matches 140; Conservative 89; Mismatches 232; Indels 139; Gaps 25;

QY 82 VLDKEQTGF-----LAEPADPNVTKLRSGYFSSKVSLSL---TEKDG-AYTWHITPGPRTK 133
DB 79 VVGKEQGEFRLEYLLSQAEFTQRALEPFQGYAPTIRIDAPRQNDHITVVIYVDKGEFVR 138
QY 134 IANVGVAIIGDILSDGNLAERYNALENMQOPVGSDFDQDSWENSKTSVLGAVTRKGVPL 193
DB 139 VQRAHVAMTGAAGD----HYLQDLEDFPKLGEIFNIPTVEASKVITRRLAERGTFD 194
QY 194 AKLGNTRAAVNPDATVDLNVVDSGR-----PIAFG-DFEITGTQRYPEOIVSGLARFQ 247
DB 195 ADFTHRREYVTRAAHAADIDLWESGRDYDMGVRFRHYD-----FHEGLFNPLVYWD 247
QY 248 PCTPYDLDLLDFOALEONGHYSASGASVQADPRLQGD-RVPVKVSVTEVARKHLEGTIR 306
DB 248 EGSYFHEGKLDRLRESLTKLDYSSIDIQPKPEADPEGVPVDVKLERASKIYTAGIS 307
QY 307 LDSEYGLGGKIADYYNLFNPKNGYIGSVWDMDKYETTLAAGISQP-RNRYGNWYTSNVSY 365
DB 308 YGSESGAG-----LRAVERMYNARGHKHNAALDY 338
QY 366 NRSTQNLKRAFSGGI-----WVRDRAGIDARLGAELPAGRKIPGSDIDLG--- 414
DB 339 ----AQNLSLTATYAOIPAFKLDGWYI-----FFRAYDEQTKYIDLNRVK 381
QY 415 -----NSHAT-----MLTASWKRL-----LNNVLHPE-NGHYLDGKIGT 448
DB 382 LSAARSGOINRHLTATASNALRGRWYRADDGNTVAVQOSTLVYPQLEASYVDVDDAT 441

```

```

QY 449 TLTGFLSTALIRTSARA---GYFFTPENKKLCTF-----IIRQAGYTVARDNAD 496
DB 442 FPRNGSAATVLLRGASALGSKNFQTLHQQLRWFHGLGASSRLILRGEAGTTWASDLVA 501
QY 497 VPSGLMFRSGGASVRCGYELDSIG--LAGPNSVLPERALLVGSLEYQL-----PFRTLL 549
DB 502 MPPSLRRFPAGGVNSIRCYAFREVGPRTAKPDAPALCAKAKHVSAGAEFEYVYKGGPF---- 557
QY 550 SCAPVPHMDGDAANFKMKLKH-GSLGVNRPSPPLAFSFDIAYG--HSDKKIRWHISLG 606
DB 558 GGAVPVDGSGA---FNRYPDWHTGVGILGRYSRPGVPRVDIARGNSPDSKIOLYIDIG 614

RESULT 13
QX3V8
ID Q9X3V8 PRELIMINARY; PRT; 738 AA.
AC Q9X3V8
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 81.1 kDa protein.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Lee H.J., Kang H.S.;
RT "Sequence analysis of 42C11 fosmid clone of Zymomonas mobilis ZM4.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF088896; AAD21534.1; -
DR InterPro: IPR000184; Bac_surfaG_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Hypothetical protein.
SQ SEQUENCE 738 AA; 81140 MW; F169545D09168114 CRC64;

Query Match 9.5%; Score 300; DB 2; Length 738;
Best Local Similarity 23.9%; Pred. No. 1.1e-10;
Matches 165; Conservative 92; Mismatches 264; Indels 168; Gaps 33;

QY 24 ADLSENKAAGFALPKKSPDTSVKLKPRFPVRIDTQDSEIKDMVEEHLPLITQOEE--- 80
DB 115 SDASEKTA-----QNIETTSVTLDPALSEPLDLSHFMDM-NEHVLAKESDDDLRY 167
QY 81 -----EVLDEQET-----GFLAEAPDNVTKMLRSKYFSKY 113
DB 168 KVDVKGLSNIDTQAFKNYSVLLSNKNKADSLSVIGARATNDQELINILRSQGYDGA 227
QY 114 SLT----EKDGAYTVHTPGPRTKIANVG-VAILGDLSDGNLAERYNALENMQOPVGS 168
DB 228 ALSITPEK-GQYVRYD-----VNAGSIYKLGQINLTGSEEPRLIARVALNLHKG 279
QY 169 FDOODSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDATVDLNVVDSGRPIAFGD 228
DB 280 PLIATHIEQAENNLITLPHYGYPFAKIGDRTILLDDETHTGDTYTLPNVAGNVSGYSII 339
QY 229 ITGTQR--YPEOIVSGLARFQPGTPYDLDLLDFOALEONGHYSASVQ----- 276
DB 340 VSNKHIVLDAKHLISHIARFKORYDSQWDDLRQALAAATLSFHSVSEPIATGRKRD 399
QY 277 ----ADFRLQGDVRPVKVSVEVTKRHKLETGIRLDSYGL-----GGKIADYY--- 322
DB 400 GSEIVDLVDRG-----RGKKH-----STAVTGGYGTGEGFKAQGSWLSRNPPE 445
QY 323 -----NLFNKCYIGSVVMDMK-YETTLAAGISQPRNYGNWYTSNV 364
DB 446 GSLTFSGILGTROOQLSALFNRNAGA---RDRVLIQTLTACRERYDAYNGYFSLGGS 501
QY 365 YNRSTQNLKRAFSGGIYVWRDAGIDARLGAELPAGRKIPGSDIDLGNHSHAT---ML 421
DB 502 LSROSTOLWOKR-----W-----TYSIGAEILTQTNER-----SYDFSRSERLTFYL 543

```

[illegible]

```
Db 341 NSDPKNYYLLGASFAKPGVYPTDQFVATLDKAREVLDAYTETSINAKTGFQTQFSDE 400
Qy 372 -----NLEKRAFSGGIWYVD--RAGIDARLGAELAEGRKIPGSDIDLGNHSHATMLT 422
Db 401 LSGALYANASQGHFVDDVFGKRDEFTAGLEGNL-----LYDSR----- 438
Qy 423 ASMKROLLANVLHPENGHVLDCKIGT-----TLGTFLSSTALIRTSARAGYFFTPENKKLG 478
Db 439 -----NNKPDPSGFLVGNITOPFVEFHYGNFATR-----FTRAGRTYH 477
Qy 479 TF-----IIRG--QAGYTVARNADNPVSGLMFRSGGASSVRGYELDSIGLAGPNSGVLP 530
Db 478 GFQOTDRVVYLAGRLKVGSTGVGSIADLPSPQLFLAGGSGSVRGYGVYRIGVSAGNIIIG 537
Qy 531 ERALLVGSLEYOLPFTRTLSCGAVFHDG-----DAAANFKRMKLGKHSGLGVRFESPLAPP 586
Db 538 GRSLSVANGEVTRITDSIGAVAFDAGYVGEKSFDPFSE-QMRVGVGGGLRLTSLGPI 596
Qy 587 SFDIAY-----GHSDDKKIRMHISLGRPF 609
Db 597 RLDVAVPLNRRSGDPNFGYFVGIGQAF 623

RESULT 16
Q8YUR6
ID Q8YUR6 PRELIMINARY: PRT: 833 AA.
AC Q8YUR6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Hypothetical protein Alr2269.
GN ALR2269.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Keneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriquchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73968.1;
DR InterPro; IPR00184; Bac_surfAq_D15.
DR Pfam; PF01103; Bac_surface.Aq.2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 833 AA, 89633 MW, 7F52B1E52F8A1BE9 CRC64;

Query Match 8.8%; Score 279; DB 16; Length 833;
Best Local Similarity 22.9%; Pred. No. 3e-09;
Matches 153; Conservative 82; Mismatches 234; Indels 200; Gaps 26;

Qy 81 EVLDKEOTGLAEAPDNVKTMLRSK-----GYESSKYSLTEKDG 120
Db 225 EVLVRQSQQLTPELTQVYVNIPTQPTTTRSQLQEDINAFGTGFFSN-VQASPEDT 283
Qy 121 AYTVHIT----PGP----RTKI--ANVGV-----AIIIGDILSDGNLAERYR 156
Db 284 PLGVRYSFIVQPNVLSKVEIQANFGTVNPSVLPQATADEIFRAQYKILNLRDLOEGIK 343
Qy 157 NALENWQVPVGSDFDQDSWENSKTSVLGAVTRKGYPLAK-LGNTRAAVNPDDTATVDLNV- 214
Db 344 ELTKRYQD-----QGYVLNVVGAPOVSEN---CVVTQLVA 376
Qy 215 -----VVDGSRPIAFGEFTGTQYPEQIVSGLARPOPTGTVDLGL 256
Db 377 EGVVENISVFRNKEGQDVNEQGPi-----RGRTQDYIITREVLKPGQVFNRT 427
```

```
Qy 257 LLDPQALBEONCHYSGASVOADFRLQGRVPVVKVSTEVYKRHKLETGIRLDSYGLGCK 316
Db 428 VQKDLQRFVGTGLFEDVNVSLD-PCTDPTKVNVMVYVSSGSIAGAGISSSSGLFGT 486
Qy 317 IAYDYTNL-----FNKGYIGSVVWMDKYETTLAAGISQPRN 353
Db 487 VSYQOQNLNCRNOKLGAEVOLGERELLFDLRFDTDWIGG-----DPYRTSYTANIPRRS 541
Qy 354 YR-----GNYWTSN-----VSYNRSTTONLEKRA---FSGGI 382
Db 542 ISLIFGCKDEDIHTFDGPNNDTNGQDRPRVTRLGGGVTFTRPLSANPFERAEMTASAGL 601
Qy 383 WY-----VRDRAGIDARLGAELAEGRK-----IPGSDIDLGNHSHATMLTASMKROLLNNVL 434
Db 602 QYQVSTRDADGNLRKDGAVFDDNGNRTSEIVPLSFSGCTGEDDLLVLQGAQRDLRNPL 661
Qy 435 HPENGHVLDCKIGTTLTGTFLSSTALIRTSARAGYF-----FTPENKKLGTFTIRGOAGY 488
Db 662 OPTSGSFL--RFGVDSQSVPGSGNIFLTRFRGSYSQYLPVKPTGFGSGPETIAFNIOGGT 719
Qy 489 TVARNADNPVSGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERALLVGLEYQLPFTRT 548
Db 720 VL-----GDLPPYEAFTLGGSNSVRGYEGALGSG-----RSFVQASVYRFPVFSV 766
Qy 549 LSGAVFHDGCD-----AAANFKRMKLGKHSGLGVRFESPLAPPFSDIAYG-HSDKKIR 600
Db 767 VSGALFFDVGSDLGTSYTRTAEVLNKGSGGYGLGVRFVQSPGLPIRID--YGINDGDGR 834
Qy 601 WHISLGRTF 609
Db 825 INFGIGERF 833

RESULT 17
Q9ATW6
ID Q9ATW6 PRELIMINARY: PRT: 628 AA.
AC Q9ATW6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Hypothetical protein CCI1603.
GN CCI1603
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CBI5;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AC005835; AAK23582.1;
DR TIGR; CCI1603;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 628 AA, 68203 MW, 92DAD1722CF82295 CRC64;

Query Match 8.6%; Score 272.5; DB 16; Length 628;
Best Local Similarity 23.2%; Pred. No. 4.9e-09;
Matches 138; Conservative 92; Mismatches 205; Indels 81; Gaps 19;

Qy 60 ODSEIKDMVEEHPLTITQOOEEVLKDEQGTGLAEAPDNVKTMLRSKYFSKVV--SLTE 117
Db 68 EDRLATQATRAL---SDSKQPPRSRSEARRRQAGDAIAYLRAEYGYAYTPEPYTE 124
Qy 118 KDAYT-VHITPGPRTKIANVGVAITLGDITLSDGNLAERYRNALENWQVPVGSDFDQDSWE 176
```

```
Db 125 GPPRAIVRITGPAFLADPHIDKSGPPDEGVR-----QRAVAMRLTEGER 173
QY 177 NKSITVLGAVTRKGYPLAKLGNTRAAVNPDTATVD-----LNVVDSGRPIAFGDPEI 229
Db 174 GRSADVGAEGRIVAAKGLGVADYAAEPREVVDHADRTVTRPTFRIMAGELVRLNGVDV 233
QY 230 TGTORYPEQIVSGLARFPGTGYDLDLDDLLDQQALEQNGHYSGASVQ---ADFDRLQGD 286
Db 234 VTGGRTPNEMVGRGLAPWVAGDYPDEVAELERLRLDTAVYDSISVSLAGTDKASAEGR 293
QY 287 VPVKSVTEVKRHKLETGIRLDSVGLGKTIAYDYNNFNKYGISGVVWMDMKYETTLAA 346
Db 294 -PWWTLDORRARTIELGAGYSTSGAGVDARWIRYNRQRADTYYALRFAKLEORLGA 352
QY 347 GISOPNRYGNWTYS-----NVSYNRTTQNLKRAFSGGIYVVRDRAGIDAR-LGAE 398
Db 353 EISLP-----HWRRPQOTKLINSSVFRNDTDAYNKGATVGVDLTRRQTTAYRTFGVS 406
QY 399 F-----LAEGKIPGSDIDLGNSHATML-TASKKROLLNNVLPENHGYLDG 444
Db 407 FDSQTKQVNRNGLIAGKRL-----NLATLAGLAAYWDFSDDILDPKRGWRLET 457
QY 445 KIGTTTLGTLSTALIRTSARAGYFETTPENKKGITFI-IRQAGYTVARDNADVPSPGLMF 503
Db 458 RAEPYTVAGDTSVPVYKLQAG-GSAYLPFGKODSTVLAARVKLAGILGALDVPASRRF 516
QY 504 RSGGASSVRCYELDSIGLA---GPNGSVLPERALLVGSLEYQLPFTRTLSGAVFHDMDG 559
Db 517 FSGGGVSGHYAYQAYQPRLSNDTPQGGI-----SLVETSPVROKITDRWSGVAFYDAGA 572
QY 560 AANF--KRMKLHGSGGVWRVSPFAPSFEDIA-----YGHSDKKIRWHISLGTRF 609
Db 573 IGHETTPQREDFRAGLGVRYDLGFGPIRADIAAPLGRKRGDGPKEIYLSIGQSF 628
```

RESULT 18

```
P95359
ID P95359 PRELIMINARY; PRT; 792 AA.
AC P95359;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Outer membrane protein.
GN OMP85.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RA Manning D.S., Reschke D.K., Judd R.C.;
RT *omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis are similar to Haemophilus influenzae D-15-Ag and Pasteurella multocida Oma87.;
RL Microb. Pathog. 23:0-0(1998).
DR EMBL: U81959; AAC17600.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
SQ SEQUENCE 792 AA; 87868 MW; 90E32D24AA0513D8 CRC64;
```

Query Match 8.2%; Score 260.5; DB 2; Length 792;
Best Local Similarity 22.6%; Pred. No. 4.1e-08;
Matches 139; Conservative 86; Mismatches 242; Indels 149; Gaps 28;

```
QY 97 DNKVTMLRSGYF-----SSKVSLETKDCAYTVHT--PGPRTKIANVGVAILDILSDG 149
Db 223 EKVTDFYQNGYFDFRLDLDIQTNEHTRTQITIKTVHSGRFRWGVSI-----EG 274
QY 150 NLAEYTRNALE-----NMOPVGSDFDQDSWENSKTSVLGAVTRK-----GYPLAKTG 197
Db 275 DTNEVPKAELEKLLTMKPGKWTY-----RQOMTAVLGEIQNRMGSGAYGYSI- 322
```

```
QY 198 NTRAAVNPDTATVDLNVVDSGRPIAFGDPEITGTORYPEQIVSGLARFPGTGYDLDLL 257
Db 323 SVQPLNAGTKTVDFVHLHLEPGRKIYVNEIHITGNKNKTRDEVRELRQMESAPYDTSKL 392
QY 258 LDFQQALEQNGHYSGASVQADFRLQ--DRVVKVSVTEYKRHKLETGIRLDSVGLG 315
Db 383 QRSKERVELIGYFD--NVQFADVPLAGTPDKVDLNMSLTERSTGSLDSAGWQDVTGLVM 440
QY 316 KIAIDYNNLPNKGYSIGSVVWMDMKYETTLAAGISOPNRYGNWTYS---NVSYN----- 366
Db 441 SAGVSDQNLFGTG--KSAALRASRSKTTLNGSLT-----FTDPYFTADVSLGIDITYKAF 494
QY 367 ----RSTTONLEKRAFSGG-----TWYVRDRAGIDA-----RLGAEPFLAE 402
Db 495 DPRKASTSVQVKYKTTTAGGVRNGIPVTEYDRVNFGLAAEHLTVNTYNKAPKRYADFIKQ 554
QY 403 GRKIPGSDIDLGNSHATML---TASKKROLLNNVLPENHGYLDGKIGTTLGTLFSLSTAL 459
Db 555 YGKTDGAD---GSFKGLLYKTVGWGRNKTDSALWPRG-YLTG-----VNAEI 599
QY 460 IRTSARAGYFETTPENK-----KLGTFIIRQAGYTVARDNA-DVPFGLMFRSGGASSV 511
Db 600 ALPGSKLQIYVSATHNOTWFFPLSKTFTLMLGCEVGIAGYGRTRKEIPFENFYGGILGSV 659
QY 512 RGVELDSIGLAGPN-----GSVLP---ERALLVGSLEYQLP---FRTLSGAVFHDMDG 559
Db 660 RGYE---SGTLGPKVYDEYGEKISYGGNKKANVSAELFPMPPGAKDARTVRLSLPADAGS 716
QY 560 -----AANFKRM-----KLHGSGGVWRVSPFAPSFEDIA-----G 593
Db 717 VMDGRYVTAENGNNKSVYSENAHKSTETNELRYSGAGVTVLSPGLPMKFIAYPLKKK 776
QY 594 HSDKKIRWHISLGTRF 609
Db 777 PEDEIQRFQFOLGTTTF 792
```

RESULT 19

```
Q9K1H0
ID Q9K1H0 PRELIMINARY; PRT; 797 AA.
AC Q9K1H0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Outer membrane protein Omp85.
GN NMB0182.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT *Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.;
RL Science 287:1809-1815(2000).
DR EMBL: AE002375; AAF40639.1; -
DR TIGR: NMB0182; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
RW Complete proteome.
SQ SEQUENCE 797 AA; 88436 MW; B68BFC5A43D22EB8 CRC64;
```

Query Match 8.0%; Score 253; DB 16; Length 797;
Best Local Similarity 21.9%; Pred. No. 1.3e-07;

Matches	135;	Conservative	89;	Mismatches	248;	Indels	144;	Gaps	26;
Qy	97	DNYKTMRLRSKGYF-----SSKVSYTEKDGAVTVHIT--PGPRTKIANVGVAILGDILSDG	149						
Db	223	EKYTFYQNGGYFDFRILDTIOTNEDKTKITIVHEGGRFRWGVSI-----EG	274						
Qy	150	NLAEYYRNALE-----NNOOPVGSDFDSDSENSKTSVLGAVTRK-----CYPLAKLG	197						
Db	275	DTNEVPKAELEKLLTMKPKWYE-----RQQTAVLGEIQNRWMSGAGYAYSEI-	322						
Qy	198	NTRAAVNPDTATVDLNVVDSGRPIAFGDEITQYREQYISGLARFQGPPTVDLDLL	257						
Db	323	SVQPLPAETKTVYDVLVHLIEFGRKLYVNEIHTGNKTRDEVVYRRELROMESAPYDTSKL	382						
Qy	258	LDFQQALEQNGHYSGASVQADFRLQG--DRVVPKVSVEVKKRKLLETGIRLDSYELGG	315						
Db	383	QRSKERVALLGYFD--NVQFQAVPLAGTPDKVDLNNLSLTERSTGSLDLSAGWVQDTGLVM	440						
Qy	316	KIAYDYNLFNKGYGISVVVMDKYEYETL-----AAGISOPRNYRGNW----	359						
Db	441	SAGVSQDNLEFGTG--KSAALRASRSKTTLNGLSLFTDPYFADGVSLGYDVYKAFDPKR	498						
Qy	360	-TSNVSYNRESTTONLEKRAFSGGIWVDRAGIDA-----RLCAEFLABGRK1	406						
Db	499	ASTSIQYKTTTAGAGIRMSVPVTEYDRVNFGLVAEHLTYNTYKAPKHYADFIKKYCKT	558						
Qy	407	PGSDIDLGNSHATML--TASWKRQLLNNLHPENGHYLDGKIQTTLGTLSLSTALIRSPA	464						
Db	559	DGTD--GSGFKWLYKGTGVGRNKTDSALWPTRG-YLTG-----VNAETALPGS	604						
Qy	465	RAGYFFTPENK-----KLGTFIIRGOAGYTVARDNA-DVPSGLMFRSGGASVNGVEL	516						
Db	605	KLOYYSATHINQWTFPLSKLFTMLGGEVGLAGCYGRTEKPEIFENFYGGGLGSVRGYE-	663						
Qy	517	DSGLAGPN-----GSVLP---ERALLVGSLEYQLP---FTTTLGSAVFHMGD----	559						
Db	664	--SGTLGPVKYDEYGEKISYCGNKANVSAAELFPMFGAKDARTVLRLSFADAGSYWDCK	721						
Qy	560	-----AAANFRM-----KLKHGSLGVWRFSPPLAFPSFDIAY-----G	593						
Db	722	TYDDNSSATGCRVQNIYAGNTHKSTFTNELRYSAGGAVTWLSPGLPMKFSVAYPLKKK	781						
Qy	594	HSOKKIRWHISLGRF	609						
Db	782	PEDEIQRFQFGTGF	797						
RESULT	20								
Q9JX31									
ID	Q9JX31	PRELIMINARY;	PRT;	797	AA.				
AC	Q9JX31:								
DT	01-OCT-2000 (TREMBLrel. 15, Created)								
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)								
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)								
DE	Outer membrane protein OMP85.								
GN	OMP85 OR NMA0085.								
OS	Neisseria meningitidis (serogroup A).								
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.								
OX	NCBI_TaxID=65699;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;								
RX	MEDLINE=20222556; PubMed=10761919;								
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,								
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,								
RA	Davies R.M., Davits P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,								
RA	Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,								
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,								
RA	Whitehead S., Spratt B.G., Barrall B.G.;								
RT	"Complete DNA sequence of a serogroup A strain of Neisseria								
RT	meningitidis 22491."								
DR	Nature 404:502-506(2000).								
DR	ENBL; AL162752; CAB83401.1								

DR	InterPro: IPR000184; Bac_surfAg_D15.
DR	Fram: Pf01103; Bac_surface_Ag; 1.
KW	Complete proteome.
SQ	SEQUENCE 797 AA; 88404 MW; 65D4E7E0C9E1D1F CRC64;

Query Match	
Best Local Similarity 21.9%; Score 253; DB 16; Length 797;	
Matches 135; Conservative	89; Mismatches 248; Indels 144; Gaps

Qy	97	DNVKTMIRSKGYF-----SSKVSLTEKXDGAYTVVHIT--PCPRTKIANYGVAVILGDIISDC 149
Db	223	EKYTDYQNNGYDFRILLDTIQINEOKTKQTIKITVHBGGHRMCKYSI-----EG 274
	:	: :
Qy	150	NLAELYRNALE-----NNQPVGSDFPDQSWENSKTSVLCAVTRK----GYPLAKLG 197
Db	275	DTNEVPKALEKLITMKPGWE-----ROOMTAVLGEIQNMWSAGSAYVEI- 322
	:	: :
Qy	198	NTRAAVNPDATLVUNVVDSGRPIAFGEDEITGTORYPEQIVSGLARFPQPTVDLDLL 257
Db	323	SVOLPNACITKTVDPLVHLIEPGRKIYYNEITHITGNKNKTREYVRRELROMESAPYDTSKL 352
	:	: :
Qy	258	LDPQQALEQNGHYSGASQAQDFRLQG--DRVPPVKVSVTEVKRHKLETGITRLDSEYGLG 315
Db	383	QRSKERVELLYGFD--NVQPDVPLACTPKVDLNLSLTERSTGSLDSLARGWQDTGLVM 440
	:	: :
Qy	316	KIADYINLFNKGYIGSVVDMNDXYETTL-----AAGISQPRNRYGNW--- 359
Db	441	SAGVSODNLFGTG--KSAALRASRSKTTLNGSLSTDPDYFTADGVSGLGDYVGKAFDPRK 498
	:	: :
Qy	360	-TSNSYNRSTTNLEKRAFSGGTWYVRDAGIDA-----RLGAEPFLAEGRKI 406
Db	499	ASTSIQKQYTTTAGAIRMSVPVTEYDRVNFGLVAEHLTVNTYNKAPHYADFKKYKGT 558
	:	: :
Qy	407	PGSDIDLGNSHATML--TASHKRROLLNNVLIPENGHYLDGKLTGTLTFLSSTALIHTSA 464
Db	559	DGTD--GSFKGWLYKGVTGWRGNKNTDSALWPTRG-YLTG-----VNASIALPGS 604
	:	: :
Qy	465	RAGYFFTPENK-----KLGTFTIRQAGYTVARDNA-DVPSGLMFRSGASSVRGYEL 516
Db	605	KLOYYSATHNOTWFPPFSKTFTLMLGCEVGIAGGYGRTWKEIPFENFYGGGLGSVRCYE- 663
	:	: :
Qy	517	DSITGLAGPN-----GSVLP----ERALLVGSLEYQLP---FTRTLSGAVRFHMDGD----- 559
Db	664	--SGTLGPVYDEYGEKISYCGNKANVASLELUPMPGAKDARTVRLSLPADAGSVDWGK 721
	:	: :
Qy	560	-----AAAANKRM-----KLKHGSLGLVRFSPFLAPSFDIAY-----G 593
Db	722	TYDNSSSATGCRVQNIYGACNTHKSTFTNELRYSAGGAVTWLSPDKPMKFSYAPLUKKK 781
	:	: :
Qy	594	HSDKIRWHISLGRFR 609
Db	782	PEDEIORFQQLGTTTF 797
	:	: :

RESULT 21	
O30912	PRELIMINARY; PRT; 797 AA.
ID	O30912:
AC	O30912:
DT	01-JAN-1998 (TREMBLrel_05, Created)
DT	01-JAN-1998 (TREMBLrel_05, Last sequence update)
DE	01-DEC-2001 (TREMBLrel_19, Last annotation update)
DE	Outer membrane protein Omp85.
GN	Omp85.
OS	Neisseria meningitidis.
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX	NCBI_TaxID=487;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=HH;
RA	Manning D.S., Reschke D.K., Judd R.C.;
RT	"Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis
RT	are similar to Haemophilus influenzae D-15-Ag and Pasteurella
RT	multocida Oma87.";

```

RA Microb. Pathog. 23:0-0(1998).
DR EMBL: AF021245; AAC17599.1;
DR InterPro: IPR000184; Bac_surAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
SQ SEQUENCE 797 AA; 88539 MW; CF911B5F70B9999CF CRC64;

Query Match 7.9%; Score 251; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 1.7e-07;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

QY 97 DNWKTMLRSKGYF-----SSKVSUTKDCGATVTHIT--PGPRKIANVGVAILDLSLG 149
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 223 EKVTDPYQNGYDFRILDTDIQTNEDTKTKITVHEGGRFRWKGKVSF-----EG 274
QY 150 NLAEYRNAL-----NMQOPVGSDFDQSWENSKTSVLGAVTRK-----GYPLAKLG 197
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 275 DTNEVPKAELEKLLTMKPGKVE-----RQMTAVLGEIQNRMGSGAVYSEI- 322
QY 198 NTRAAVNPOTAVDLNVVDSRPIARCDPEITGTQRYPEQIVSGLARQPGTPYDLDLL 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 323 SVQPLNAETKTVDVFLHIEFGKYYVNEIHITGNNTREDEVVRRELROMESAPYDTSKL 382
QY 258 LDFQQALEONGHSGASQAQDFRLQG--DRVPVKVSVTEVRKHLETGIRLDSSEYGLGG 315
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 383 QRSKERVLLGYFD--NVQDAVPLAGTPDKVDLNMSLTERSTGSLDSAGWYQDTGLVM 440
QY 316 KIAYDYNLFNKGYYGVSVDMDKYEYTL-----AAGISOPRNYRGNYW---- 359
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 441 SAGVSODNLFUGTG--KSAALRASRSKTTLNGSLSTFDPTADGVSLGDYDVYKAFDPRK 498
QY 360 -TSVNSYRSTONLEKRAFSGGIWYVRDRAGIDA-----RLGAEPLAEGRI 406
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 499 ASTSIKOYKTTTAGIRMSVPVTEYDRVNFGLVAELHTVNTYKAPKHVADIKKYGT 558
QY 407 PGSDIDLGNHSHATML--TASWKROLLNNVLPENGHYLDGKIGTTLTGLTSALRTSA 464
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 559 DGYD--GSGFKGWLYGKGVGCRNKTDSALMPTRG-YLTG-----VNAEIALPGS 604
QY 465 RAGYFTTPENK-----KLATFIIRGQAGYTVARDNA-DVPSGLMFRSGGASVVRGYEL 516
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 605 KLOYSATINQWTFPLSKTFTLMLGGEVGIAGGYGRTKBEIPFENFYGGSLGVSRYE- 663
QY 517 DSIGLAGPN-----GSVLP---ERALLVGSLEYQLP---FRTLSCAVFHDMDG----- 559
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 664 --SGTLGPKVDEYGEKISYGNKKNVSAELFPMGAKDARTVRLSLFADAGSVDGK 721
QY 560 -----AAANFKRM-----KLKHGSLGVRFSPFAPSFDTAY----G 593
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 722 TYDNDSSATGCRVQNIYAGNTHKSTFTNELRYSGAGVATWLSPLGPMKERVAYPLKKK 781
QY 594 HSDKKIRWHISLGTRE 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 782 PEDEIORFOFQLGTTTF 797

RESULT 22
O67326 PRELIMINARY; PRT; 778 AA.
ID O67326
AC O67326;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane protein.
GN OMP OR AQ_1300.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL Nature 392:353-358(1998).
DR EMBL: AF000733; AAC07292.1;
DR InterPro: IPR000184; Bac_surAg_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 778 AA; 89998 MW; 9C1C1948DB6C93A CRC64;

Query Match 7.8%; Score 246; DB 16; Length 778;
Best Local Similarity 22.3%; Pred. No. 3.4e-07;
Matches 137; Conservative 88; Mismatches 270; Indels 120; Gaps 25;

QY 61 DSIKDKWVEEHLPLITQOQEEVLDEKQGTFLAEEAPDNVKTM---LRSGYFSSKVSLTE 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 218 DRKLKGLMELKSPNI-----FLRLHAPAFSEETLKEDYKKEEFYKNEGFLVEVSEY 271
QY 118 KOGA----TVTHITPGPRKIANVGVAILDLSGDLAEYRNALN---WQOPVGSDF 170
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 272 KEGPARKVIVVKEGPRYKLEIKI-----EGNTLFAYSELVDNLKKRKGRYF 323
QY 171 DQDSWENSKTSVLGAVTRKGYPLAKLG--NT---RAAVNPDTATVNLVWVDSGRPIAF 224
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 324 RREVIKINRI-----REKY--AEIGFLNTSVEERVNVNDEKKEVSVLLKIEGRPVY 376
QY 225 GDEITGTQRYPEQIVSGLARQPGTPYDOLLDFQOALBN-----GHYSGASVQ- 276
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 377 KKIKRGNTETRDYVIRREMYOENE-----LALKGIERSKTRIMNIGYEDVQIEP 429
QY 277 -----ADFORLQDRVPVKVSVTEVRKHLETGIRLDSSEYGLGKIAYDYVNLFNKGYIG 331
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 430 VPRDADWDL-----VKIREFTQGSVGLSNEVTLGFIELRKGNGRGTGDI- 480
QY 332 SVVWDMKTYETTLAIGISOPRNYRGNYWTSVSNR-----STTONLEKRAFSGGI---WY 384
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 481 -----AGISV--SYGLYRNNAISYTRKWFLKKPVLDLSAFDRRIEYDY 524
QY 385 VRDRAGIDARLGAELAEGR-----KIPGSDID-----LCNSHATMLTASWK 426
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 525 TVERTGFSVALSKELSEYWRASIGTSIORVKYSDIDPASPYYVKEQAGRRSRKIFTLT 584
QY 427 RQLNNVLPENGHYLDGKIGTTLTGLTSALRTISARAGYFTTPENKKLGTGFI-IRQ 485
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 585 RDRTRDYLLPTKGSLSVFRNVSGLLGGDEKFKYFEVGEAKYFSDTYFTDGIILSLKE 644
QY 486 AGYTVARDNADVPISGLMFRSGGASVVRGYELDSIGLAGPNC-SVLPERALLVGSLEYQLP 544
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 645 VGFVEGYGKKVPIDERRFPVGGDSIRGYEYAGVPDPTLDPIGAKKKVITSELMYP 704
QY 545 -FRTLSCAVFHDMDG---DAAANFKRMKLGKSGLVGNWPSPLAPSFDTAY-----GH 594
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 705 VFKKMLYFAGFDYGLGADKWSDFKLSNPRGSGYGVIRIITPPAPRIDWAFKTKTVPGD 764
QY 595 SDDKKIRWHISLGTRE 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 765 TNRS-RIHFVLGTTF 778

RESULT 23
Q82H58 PRELIMINARY; PRT; 795 AA.
ID Q82H58
AC Q82H58;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative surface antigen.
GN YPO1052.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
```

```
RC SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Partridge M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K.P., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltham T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT *Genome sequence of Yersinia pestis, the causative agent of plague.*;
RL Nature 413:523-527(2001).
DR EMBL; AJ414146; CAC89894.1; -.
DR InterPro; IPR000184; Bac_surfaq_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 795 AA; 87836 MW; 8C4581328D38F821 CRC64;

Query Match 7.8%; Score 246; DB 16; Length 795;
Best Local Similarity 21.2%; Pred. No. 3.5e-07;
Matches 142; Conservative 105; Mismatches 273; Indels 150; Gaps 30;

QY 47 VKLPKFPVRIDTQDSIKDMVEHLPLLTQQOEEVLQKEQTFGLAEAP----- 96
DB 163 VDLKLVTEGVSARKIQI-NIVGNH-----SFTDELISRQ---LRDEVPMNVVGRDKY 214
QY 97 -----DNVKTMLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGPRTKTANVGVA 140
DB 215 QKQKLAGDLETLRSFYLDGRYARFNIDTQVSLTPQKGIYVITINITEGPOFK----- 267
QY 141 ILGDILSDGNLAIEYRNALNENWQPVGSDFDQDSDWENSKTSVLGAVTRKGYPLAKLGNT 200
DB 268 -LNSVIVSGNLAGHQSAAELKTIKPELFGNSKVTREDDIKKMLGRYGVAYPRV-VTQ 325
QY 201 AAVNPDTATVLDNVVDSGRPIAFGDFEITGRTQYPEQIVSCLARFQPTPYDILLDF 260
DB 326 PEINDDKTKVLHINDAGNRVYRHIREGNDTSKDSVLRREMQMAGLMDQVEAG 385
QY 261 QOALEQNGHYSGASVOADFRLQD---DRVVPKVSYTEVKRIKLETGIRLDSYGLGGKIA 318
DB 386 KERLNRLQYFE--TVDVETQVPGAADLVDTYKVKERTGSLNFGICYGTESGVSFQV 443
QY 319 YDYVNLNFKGYIGSVVMDKYEITTLAAGISOPRNVRYNYT-----SNVSYNRSTTQ 371
DB 444 VQDNWILGTGNTVGINCTKNDYQTYAEFTLMDP-----YFTVGVSLGGRIFYNDKAD 497
QY 372 NLEKRAFSGGIWYVRDAGIDARGLAEFLAERKIPGSDIDLGNSH-----ATML--TASWK 426
DB 498 NAD---LSG---VTNSSYGADGTLGPP-INENNSL---RVGVGYVHNDLSMDLPQVAMWR 547
QY 427 ROLLNNVLPENGHYLDKIGTIGTFLGTFLSSTALIRTSARAGYF----- 469
DB 548 --YLESV--GERPCY--DCRGFTTDDFTLNLGWTNNLDRGFFTSVGKSSVNTKIITVP 602
QY 470 -----FTPENKKLTFFII-RQAGYTVARNADVPVSGLMFRSGGASSVRGY 514
DB 603 SDNEFFYKVTEDTSAYQPLNEDRSWVLLGRGLGYDGIGSKEMPEYENFYAGSGSTVRGF 662
QY 515 ELDSITGLAG---PNG-----SVLPERALLVGSLEYQLP-----FTRTLSCAVFHD 556
DB 663 RSNIGPKAAYANGGATVNTSDAVGGNMAVASIELITPTPFISEKYSNVSTSFID 722
QY 557 MGDA-----AANF-----KRMKLUKSGSLGVNRWFSPLAFSPFDIAYG-----HSD 596
DB 723 SCTVNDTWNENTAKTRAAGIPDYGKASINRVSGVALQWMSPLGLPVSFSAKPVKDYEGD 782
QY 597 KKIWRHISLG 606
DB 783 KSEQFQFNIG 792

RESULT 24
Q98K30
```



```
Db 501 ADDRFGSGGSGVRGAYGGIGPKDFTCQPIGGLSPFETSVMRIAVTDTIGIVFVDAG 560
QY 559 DAAA-----NFRKMLKHGSLGVRFWFLAPFSDIA 591
Db 561 TVSTKSVNFSVDKV--GAGVGLRYVTFFGFLRIDAA 595

RESULT 25
Q82RPO PRELIMINARY; PRT; 804 AA.
AC Q82RPO:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative outer membrane antigen.
GN YAEU OR STM0224.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ARCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston K., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL: AAL19188.1; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 804 AA; 89526 MW; 6197447548PAC41E CRC64;

Query Match 7.2%; Score 228.5; DB 16; Length 804;
Best Local Similarity 20.7%; Pred. No. 4.6e-06;
Matches 139; Conservative 101; Mismatches 287; Indels 145; Gaps 27;

QY 47 VKLKPFPVRIDTQDSEIKDMVEEHLPLITQOOEEVLDRKQTCFLAEAP----- 96
Db 163 VDLKLVQEGVSAKIQI-NIVGNH-----AFSTEELISHFQ---LRDEVPWNVVGDRLKY 214
QY 97 -----DNVKTMLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGVA 140
Db 215 QKQKLAGDLETRSYLDGRYARFNIDSTQVSLTPDKKGIYITVNITEGQYKLSGVQVS 274
QY 141 ILGDILSDGNTAEYRNALNNQO-PVCSDFDQDSWENSKTSVLGAVTRKGYPLAKLGN 199
Db 275 -----GNLAGHSAE-IENLTKEPGELNYKVTKMEDD1KKLLGRGYAYPRV-OS 324
QY 200 RAAVNPDTATVDLVNVVDSGRPIAFGDFEITGQRYPEQIVSGLARFORFGTPYDLDLLID 259
Db 325 QPINDADTKVLRVNDAGNRFVYKIRFEGNDTSKDSVLRREROMEGAWIGSLVDQ 384
QY 260 FQOALFONGHYSASVQADFRLQ--DRVPKVSVEVRRHKLGTIRLDSEYGLGGKI 317
Db 385 GKRLNRLGFFB--YVDTDTRVPGSDVDVYKVKERNVTGSFNGIGYGESVGFQA 442
QY 318 AYDYNNLFNKGIGSVVMDKYEITTLAAGISOPRNYRGNWT----- 360
Db 443 GVQODNWLGTGYSVINGTKNDQYOTISELSVTNP-----YFTVDGSLGRIFYNDFOA 496
QY 361 -----SNVSNRSTQNL---EKAFSGGIWYVRDR-AGIDARLGAEEFLAEGRKIPG 408
Db 497 DDAIDSDYNTKSGTDTVLGTFPIEYNTLRAGLYGVHKLNNQPOQIAMDRLYLSM---G 553
QY 409 SDIDLGNSHATMLTAS--WKQOLLNNVLPENH--YLDGKIGTITLGTLSLTALIRTA 464
Db 554 QSADTSSFAADDTFNYGWTYNTKLDGRGVFPDGRVNLTKV-TIPG---SDNEYKYSVL 609
```

```
QY 465 RAGYFETPENKKLGTFIIRGAGTYVARDNADVPGLMFRSGGASSVRGYELDSIG--IA 522
Db 610 DTATVVPIDNDHKWVVLGRTRMGYGDGLGKREMPFYENFYAGGSSTVRGFSQNTIGPKAV 669
QY 523 GPNGS-----VLPERALLVGSLEYOLP-----FTRTLSCA 552
Db 670 YKNGAHTSWDDNDYEDTQESGCKSDDAVCGNAMAVASLEFITPTPTFISEKANSVRTS 729
QY 553 VFHDWG-----DAAANFRKMLKHGSLGVRFWFLAPFSDIAYG-----H 594
Db 730 FPMDMGTVMDTNMPDSSAPSDVPDYSDFGNIRMSGALQWMSPLGFLVFSYAQPPEKKYD 789
QY 595 SSKKIRWHISLG 606
Db 790 GDKAEOQFQFNIG 801

RESULT 26
Q9S341 PRELIMINARY; PRT; 797 AA.
AC Q9S341;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Outer membrane antigen.
GN OMA.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM;
RA Chatonnet-Marton P.I., Givaudan A., Lanolis A., Boenare N.E.;
RT "Photorhabdus luminescens genomic region homologous to 4.0 minute
RT Escherichia coli region promotes pleiotropic phenotypes.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ236920; CAB51929.1; -.
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surface_Ag_1.
SQ SEQUENCE 797 AA; 88713 MW; 39E72E7ABD9C14F3 CRC64;

Query Match 7.2%; Score 227; DB 2; Length 797;
Best Local Similarity 20.4%; Pred. No. 5.7e-06;
Matches 133; Conservative 97; Mismatches 261; Indels 162; Gaps 24;

QY 39 NKSPDPTESVKLKPFPVRIDTQDSEIKDMVEEHLPLITQOOEEVLDRKQ-TGLAEAPD 97
Db 185 NKS--FSSDELLNRFQLRDD-----VPMNLTADQYKQKLTGDL-----E 224
QY 98 NVKTMRLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGAILSDGN 150
Db 225 ALKSFYLDGRGYARFNIDSTQVSLTPDKKGIYITVNITEGQYKISGIDL-----NGN 276
QY 151 LAEYTRNALNNQOQPVCSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATV 210
Db 277 MAGYQSEITKLAIEPGSLYNGTQVTKMENDIKNLLGRGYAYPRV-WTQPEINDQDKIV 335
QY 211 DLNVVDSGRPIAFGDFEITGQRYPEQIVSGLARFORFGTPYDLDLLDFOALONGHY 270
Db 336 KLVNIDAGNRFYVYKIRFSGNDTQKDSVLRREROMERAWLGSDELKELKRLNLGYF 395
QY 271 SGASVQADFRLQ--DRVPKVSVEVRRHKLGTIRLDSEYGLGKIAYDYNNLFNKG 328
Db 396 E--TVDVEVQRIPGSDQVDVYKVKERNVTGSFNGVGFCTESVQFQICAOQDNWLGTG 453
QY 329 YIGSVVMDKYEITTLAAGISOPRNYRGNWTSN-----VSNRSTQNTLEKRAFSGG 381
Db 454 NAVGINASKNDYSTVAELSFSTDP-----YFTINGVSLGGRVFNDFRADDAE---LSG- 503
QY 382 IWVRDRAGIDARLG----- 396
```

```

Db 504 --YTNQSYGISGFLGFPINENNSLNFGLNYIHNSLSMDLPQVAMRYLRSNGEKPDLSEK 561
Qy 397 AEFLAERKRTPGSDIDLGSHATMLTASWKROLLNNVLHPENG--HYLDGKIGTTLGTFL 454
Db 562 AEFAAD-----FALTMGWYNNLDGRGFFTSQGVKSTLNGK-VTIPG--- 602
Qy 455 SSTALIRTSARAGYFFTPENKKLGTFTIRGOAGYTVARDNADVPGLMFRSGGASSVRGY 514
Db 603 SDNEFYKVTLDTSAYYPIINDRTWTLGRSLRGYDGLGGKELPFYENFYAGGSSTVRGF 662
Qy 515 ELDSIG-----LAPNSVLPERALLVGSLEYQLP-----FRTLSGANVF 554
Db 663 RSNNGPKAIIYKDGSPKESPRDAGVGNAMAVASLELITPTFFLDYSKYSNVRSTFF 722
Qy 555 HDMG-----DAAA-----NFKRMKLKHGSLGVRWFSPPLAPSFEDIA 591
Db 723 IDSGTVTDMDNSAVMKSGKIPDYKSGPNIRVSAGIALQWMSPLGLVFSYA 775

RESULT 27
Q9KPMO
ID Q9KPMO PRELIMINARY; PRT; 803 AA.
AC Q9KPMO
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Surface antigen.
GN VC2252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=566;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR 16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Gwin M.L.,
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolotava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ufferback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004297; AAF95396.1; -.
DR TIGR; VC2252; -.
DR InterPro; IPR000184; Bac_surfAg_d15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1958B3838F6C041D CRC64;

Query Match 7.28; Score 227; DB 16; Length 803;
Best Local Similarity 21.7%; Pred. No. 5.7e-08;
Matches 135; Conservative 95; Mismatches 225; Indels 168; Gaps 31;

Qy 84 DKEQTGLAEAPDNVKTMLRSKGYFSKVSILT-----EKDGAY-TVHITPGPRTKIAN 136
Db 212 DKYKQVLAGDI-EALRTYVLDRLGYLKFQVDSTQVAISPDKKGVYITLNLNEGEPTVSK 270
Qy 137 VGVAILDLSDGNLAAYRYNALENWQPVGSDFDODSWENSKTSVLCAVTRKGYPLAKL 196
Db 271 --VQFRGELM--GKEAEF--TSLIPFE--IGETYNGSAVTRLSESVKVLGSGYAPQV 322
Qy 197 GNTRAAVNPDTATVDLNNVVDSGRTAFGDFEITGTORYPEQIV-----240
Db 323 -RTIPEDEKQSVLWVHEAGKRYVRDIRFVGNNSTRDVLRREMRQMEGWSLNSKD 381
Qy 241 --SCLARFQ-----PCTPYDLDLLDFOALEON-----GHYSGASVQ 276
Db 382 IETGKTRNLRLGFEFVEVQTVRVVPSGSDQVDLVYSVKEANSQWNVFGVGTGSGVSQ 441
Qy 277 ADFDR-----LQGDVVPKVSVTEVKRHKLETGIRLDSEY-----GLGKGIAYDY 322

```

```

Db 442 VGLQDNFLGSGNRVGVNAMINDYQKN-----LTLEYRDPYNNLQGVSLGKV---FY 491
Qy 323 NLPKNGYIGSVWMDKYEYTTLAAGISQPRNRYGNYWTSNVSYNRSTQNLEKRAFSGGI 382
Db 492 NOFEASAGIVDYTNESYGTSLTWGF--PFD-----ELMRPEFGI 529
Qy 383 WYVDRAG-IDARLGAE-FLAEGRKTPGSDIDLGNSHAT---MLTASKKROLLNNVLHPE 437
Db 530 GYTHINKIGNTPYLVQENFLA---AQASNIDSGGNLLTDDPDINLSWTRNNLNNSYFPT 585
Qy 438 NGHVLGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTFTIRGOAGY-----TVA 491
Db 586 AGNHQAFYKMTVPG--SDAQYFKLQYDVQRYFPLTKKHEFTLLRLGLRGYNGYGTGDG 643
Qy 492 RDNADVPGLMFRSGGASSVRGYELDSIGLGP-----NGS-----VLPERAL 534
Db 644 KDNL-PPYENFYAGGFTSLRGFSNS---AGPKAVYRDYSGSNNGSDTATDSDVSGNMI 699
Qy 535 LVGSLEYQLP-----FRTLSGAVFHDWG---DAAANFKRMKLKHGS-----573
Db 700 ALASVELIVPTPFASBEARNQIRTSIFYDMASVMDTERDY-RGKADYGNQYVYDSDPTN 758
Qy 574 ----GLGVRWFSPPLAPSFEDIA 591
Db 759 YRSSYGVALQWVSPMGLVFSLA 781

RESULT 28
Q8Z9A3
ID Q8Z9A3 PRELIMINARY; PRT; 803 AA.
AC Q8Z9A3
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein.
GN YAEF OR STY0247.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAD08682.1; -.
DR InterPro; IPR000184; Bac_surfAg_d15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89466 MW; BBF8670070E67628 CRC64;

Query Match 7.18; Score 226; DB 16; Length 803;
Best Local Similarity 20.1%; Pred. No. 6.6e-06;
Matches 136; Conservative 102; Mismatches 284; Indels 154; Gaps 25;

Qy 47 VKLKPEFPVDRIDQSEIKDMVEEHLPLITQQQREVLQKEQGLAEAP-----96
Db 163 VDLKLVFQEGYSAKIQI-NIVGNH-----AFSTEELISHQ---LRDEVPWNVVGRKY 214
Qy 97 -----DNVTMLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGPRTKIANVGA 140
Db 215 QKQKLAGDLETLSRYLDRGYARFNIDSTQVSLTPDKKGIYITVNITEGQYKLSGVQS 274

```


RL Nucleic Acids Res. 22:1637-1639(1994).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Bendlak L., Mametlak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RP [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smba gene, a suppressor of
RT the mukB null mutant of Escherichia coli.";
RL J. Bacteriol. 174:7517-7526(1992).
RP [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Smallshaw J., Kelln R.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase.";
RL Genetics 11:59-65(1992).
RP [6]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94240115; PubMed=8193897;
RA Janosi L., Shimizu I., Kajji A.;
RT "Ribosome recycling factor (ribosome releasing factor) is essential
RT for bacterial growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).
RP [7]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91317739; PubMed=1860827;
RA Shimizu I., Kajji A.;
RT "Identification of the promoter region of the ribosome-releasing
RT factor cistron (frr).";
RL J. Bacteriol. 173:5181-5187(1991).
RP [8]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90062117; PubMed=2684966;
RA Ichikawa S., Kajji A.;
RT "Molecular cloning and expression of ribosome releasing factor.";
RL J. Biol. Chem. 264:20054-20059(1989).
RP [9]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=86008268; PubMed=2995358;
RA Icho T., Sparrow C.P., Raetz C.R.H.;
RT "Molecular cloning and sequencing of the gene for CDP-diglyceride
RT synthetase of Escherichia coli.";
RL J. Biol. Chem. 260:12078-12083(1985).
RP [10]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91123198; PubMed=1991717;
RA Hirvas L., Koski P., Vaara M.;
RT "The ompH gene of yersinia enterocolitica: cloning, sequencing,
RT expression, and comparison with known enterobacterial ompH
RT sequences.";
RL J. Bacteriol. 173:1223-1229(1991).
RP [11]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91100302; PubMed=1987124;
RA Dicker I., Seetharam S.;
RT "Cloning and nucleotide sequence of the fira gene and the fira200(ts)
RT allele from Escherichia coli.";
RL J. Bacteriol. 173:334-344(1991).
RP [12]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88058790; PubMed=2824445;
RA Crowell D., Reznikoff W., Raetz C.;
RT "Nucleotide sequence of the Escherichia coli gene for lipid A
RT disaccharide synthase.";
RL J. Bacteriol. 169:5727-5734(1987).
RP [13]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88058791; PubMed=3316192;
RA Tomaszewicz H.G., McHenry C.S.;
RT "Sequence analysis of the Escherichia coli dnaE gene.";
RL J. Bacteriol. 169:5735-5744(1987).
RP [14]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88139188; PubMed=3277952;
RA Coleman J., Raetz C.;
RT "First committed step of lipid A biosynthesis in Escherichia coli:
RT sequence of the lpxA gene.";
RL J. Bacteriol. 170:1268-1274(1988).
RP [15]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92380982; PubMed=1355089;
RA Li S., Cronan J.;
RT "The genes encoding the two carboxyltransferase subunits of
RT Escherichia coli acetyl-CoA carboxylase.";
RL J. Biol. Chem. 267:16841-16847(1992).
RP [16]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094229; PubMed=1688424;
RA Zhou Z., Syvanen M.;
RT "Identification and sequence of the drpA gene from Escherichia coli.";
RL J. Bacteriol. 172:281-286(1990).
RP [17]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90370122; PubMed=2203971;
RA Eriani G., Delarue M., Poch O., Gangloff J., Moras D.;
RT "Partition of tRNA synthetases into two classes based on mutually
RT exclusive sets of sequence motifs.";
RL Nature 347:203-206(1990).
RP [18]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Miyamoto K., Inokuchi H.;
RT "Nucleotide sequence of 5'flanking region of the ribosomal RNA gene
RT (rrnH) in E. coli.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RP [19]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93094132; PubMed=1459951;
RA Gervais F.G., Drapeau G.;
RT "Identification, cloning, and characterization of rcsF, a new
RT regulator gene for exopolysaccharide synthesis that suppresses the
RT division mutation fts284 in Escherichia coli K-12.";
RL J. Bacteriol. 174:8016-8022(1992).
RP [20]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94124004; PubMed=7904973;
RA Allikmets R., Gerrard B., Court D., Dean M.;
RT "Cloning and organization of the abc and mdl genes of Escherichia
RT coli: relationship to eukaryotic multidrug resistance.";
RL Gene 136:231-236(1993).
RP [21]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;


```

Db 596 DLS--FVKIDYR-GQVFAPLTDNY-TMRPHTELGVGDCYGTSTRLPPYENYAGGNSV 650
Qy 512 RGYELDSIG-----LAGPNSVL-----PE-----RALLVGSLE--Y 541
Db 651 RGFKDXTLGPRSTPSVARNPDGTPMKNOGPDGKRYTDPDQDPEAFGPNILITGGALLF 710
Qy 542 QLPFT--RRLSGAVFHDMDA-----AANFKRMKIKHGSGLGVRFWSPLAPE 586
Db 711 PLPFVKDQRLQTLVFWDVGSFTDCTPKTKTTTNCDDIKTNDLASSVGVGLTWTALGPL 770
Qy 587 SFDIA 591
Db 771 SFSIA 775

RESULT 32
Q8YMP0
ID Q8YMP0 PRELIMINARY; PRT; 676 AA.
AC Q8YMP0;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein Alr4893.
GN ALR4893.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yanada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanoacterium Anabaena sp. strain PCC 7120."
RL NCBI_TaxID=213(2001).
DR EMBL; AP003597; BAB76592.1;
DR InterPro; IPR000977; DNA_Ligase.
DR PROSITE; PS00697; DNA_LIGASE.A1; UNKNOWN.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 676 AA; 72411 MW; C5779633037D1223 CRC64;

Query Match 6.6%; Score 211; DB 16; Length 676;
Best Local Similarity 20.9%; Pred. No. 4.5e-05;
Matches 132; Conservative 85; Mismatches 248; Indels 166; Gaps 27;

Qy 76 TQOQEEVLD---KEQTGLAEAPDNVMTMLR-----SKGYFSKSVLTKDGAIVTH 125
Db 115 TPELQEIIRQVKTQTG-----GDTSQLQDVAAILTGLFAS-----ANVNSR 160
Qy 126 ITPGPRKIANVGAIVGLDILSDGNLAAYRYNALENKQPVGSDPDQDSWNSKTSVLGA 185
Db 161 TTPSGLNVYQVQVPIVRSLSLQALTYSAQPRFQSIQKPISEGLKQAAVQVQNW 220
Qy 186 VTRKGYPLAK-----LGNTRAA-----VNPDTATVNLVNVVDSGRPTAF 224
Db 221 YADNGYNLARVLSTPNRQGLINLVAEGLVSDIKFRFVNDGKTIDSN-----GNPV-- 273
Qy 225 GDFEITQRTQPEIVISGLARFQPGTPTDLDLLDFQOALEONGHYSASQVADFRLQ 284
Db 274 -----GRTKPDFLRQL-KLQPGQVQFENIVKQVQOOLYRTGLFQSVNVAFGADTKL 326
Qy 285 DRVPKVSVTEVRKHKLETIRLDSYGLGKGIAYDYNLFNK----- 327
Db 327 DMI---VELKENGARAINLGSYNGVDVGLMTGLNYQDNIGSKNDTLLANVGLSRTDLQ 383
Qy 328 -----CY-----IGSVNDMDKY-----ETTLAGISQPNRYR 355
Db 384 DTKFISPYROTNDRLGTVTVNAFRRRSETFDEDEIKLANGDKVREGKVGKGISLQRPID 443
```

```

Qy 356 GNYWTSNVSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAFLAEGRKIPGSDIDLGN 415
Db 444 G--WNASLGFNYSTS-----IRDRQGNITPTDAQ--GNPLSVSGTGVD--- 483
Qy 416 SHATMLTASMKRQLLNVLHPENGHYLDKIKGTTTGTFTLSSTALIRTARAGYF-FTPEN 474
Db 484 -DLTTSFSATKQDRDNPINPTQGSVV--RVSTEQSVPIQGQINSMNRLLKADYSQYVPVN 540
Qy 475 ---KKLGTFTIRCOACYTVARNADVPGLMFRSGGASSVRGYELDSITGLAGPNSVLP 530
Db 541 IFNSQTPQVAFALNVQAGTVL---GNLPPYETFNLGGSNSVRGYDAGNVGSG----- 588
Qy 531 ERALLVGSLEYQLPFTTTLGSAVFHDM-----GDAA-AFKRMKMLKKGSGL---GVRW 579
Db 589 -RSYVLAESAERYFPPIVPIGVGLFADFASDLGSGDVTLGNPAGVRGKPGSGFGYCGAGIRV 647
Qy 580 FSLAPFSFDIAYGHSDK-KIRWHISLGTFR 609
Db 648 DSPGLGLRAD--YGINDQGESRVHLGIGQRF 676

RESULT 33
Q9A711
ID Q9A711 PRELIMINARY; PRT; 769 AA.
AC Q9A711;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Outer membrane protein.
GN CC1915.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
RA DeRoy R.T., Dodson R.J., Durkin A.S., Gwinn M.T., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005865; AAK23890.1;
DR TIGR; CC1915;
DR InterPro; IPR000184; Bac_surfaG_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 769 AA; 85938 MW; 7B12A1FEB6226F2B CRC64;

Query Match 6.6%; Score 210.5; DB 16; Length 769;
Best Local Similarity 22.1%; Pred. No. 6e-05;
Matches 145; Conservative 82; Mismatches 247; Indels 183; Gaps 29;

Qy 61 DSEIKDMVEEHLPLITQOQEEVLDKEQTGLAEAPDN-----VKTMLRSKY 108
Db 187 DNDLRDV-----IVTKESRWYKILTSNDNDYDPRIEYDRELKHYRNKG 232
Qy 109 FSSKY--SLTE-----KDG-AYTWHITPGPRTKIANVGAIVGLDILSDGNLAAYRYNALEN 161
Db 233 FDFRIVSSVAELADPKNGFAVYTTLEEGPKYFKGTIVTELKKL-DGNLLAQI----- 285
Qy 162 WQQPV--GSDPDQDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVNLVNVVDSG 219
Db 286 --LPVRTQLYEDERIEQATDALTFAAGAAGFAFVDV-RPRVVPNRETKTVDVVOYVREG 342
Qy 220 RPIAFGFEITGTQRYPEQIVISGLARFQPGTPTDLDLLDFQOALEONGHYSASQVADF 279
Db 343 PRVYVDRIDIVGNRTLDYVLRRLEVAEGDAYNRVLDVRSKNMRRLLGFFKEVEIE-DA 401
```

```
Qy 280 DRLOGRVPVKVSTEVKHKLETGIRLDSEYGLGGKIADYDYNLFNKGYIGSVWDMK 339
Dy 402 PGSAPORTSLRVKVEQPTGELS-----FSAGYS-----SIDK 434
Qy 340 YETTLAAGISOPNRGNWYTSNVSYNSTTQNLKRAFSGGIWYDRAGIDAR--LG-- 396
Dy 435 --LVLDVGITE--RNFRGR-----CQNLRARASVGLRQIQIDGSEPRELGRN 479
Qy 397 -----AFLAERKKIPGSDIDLG-----NSHATMLTA 423
Dy 480 LVAGVNLVTRYDLFEAFAYDTKSVGGDVRFGFLPLTNDSSMSURYTVRQDEVSAVADSLCA 539
Qy 424 SWKROLNLNVLPHNGHYLDGKIG-----TTLGTFL-----SSTALI 460
Dy 540 SGS---VSQILCLRGAYITSLIGYGLRDKRNDP INPTRGWFADLNQDLAGVGDKYL 596
Qy 461 RTSARAGYF--FTPENKLGITPIIRGOAGYTVARDNADVPGLMFRSGGASSVGYELDS 518
Dy 597 KTEADAGWYWGFT---KDL-VFSATGSPGYIEGNGGDNVRINDRFYRGG-TSPRGFEIAG 651
Qy 519 IGLAGPNSVLPERALLVGSLEYQ-----LPFTRTILSGAVFHDMDGAA----- 561
Dy 652 IGRPDISSFSNSMGAKLYAISTFELTVPTFPEQYGIKAALFSDVGTAGLLDDVDQRSP 711
Qy 562 ----ANFK-RMKLKHGSLGVWFPSLAPPSFDIAYGHS-----DKLRWHISLGRF 609
Dy 712 GVFDPNKIDNLGRASGISIDMKSPMPIRFDISRLSKEDYDRTTFRSTSTR 768

RESULT 34
O9PI28 PRELIMINARY; PRT: 739 AA.
AC O9PI28;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Outer membrane protein.
GN CJ0129C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter
OC NCBI_TaxID=197;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Peltwell T., Holtroft S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
DR ENBL: AL139074; CAB72613.1;
DR InterPro: IPR000184; Bac_surfag_d15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 739 AA; 83197 MW; B3C1891A8AA781BF CRC64;

Query Match 6.5%; Score 206; DB 16; Length 739;
Best Local Similarity 18.3%; Pred. No. 0.00011;
Matches 140; Conservative 102; Mismatches 312; Indels 210; Gaps 24;

Qy 19 AYAPAADLSENKAAGFALFNKSP-DTESVCLKPKF-PVRIDT----- 59
Dy 13 ANAATAIKDIKFTIGNLHLSNYSAINIAGLIGEEINPAKNTAILNLNYKONFENTAVE 72
Qy 60 QDSSEIKDVEHLPLIT-----QQQEEVLDEKQTCFLAE-----EAPDNVKTML 103
Dy 73 NNNGILIIIVTEPTIAKVTITIGIASNDRKQVESILGTRKGTLLDEGNKEATRIKAY 132
```

```
Qy 104 RSKGYESSKVSLEKDGAYTVHITPGPRTK-IANVGVAILGDILSDGNLAEEYRNALENW 162
Dy 133 EAKSYEDTIVEYKK-----TLENTDGLLEFIVNRGENIINDVHLSGAKKFSYSDIEPA 188
Qy 163 QQPVGSDFDQDSW-----ENSKTSLVGAIVTRKGYKPLAKLGNTRAAVNPDTATV 210
Dy 189 VVNKEFEGMGMWRNDGKLVFELSNDSSRIADYEMKKGVLQVQSSPYLKYTYDTYQA 248
Qy 211 DLNVVDSGRPTAFGDPEITG----- 231
Dy 249 NLTYFIEKGPYKIKSISTENPLFDDKQNAQTVKDLRSSAGKTINIEDIRKDKWTIETOS 308
Qy 232 -----TQRYPE-----QIVSGLAR-----FQ 247
Dy 309 ADLGYAFVEYDPDIQNDQTEATVVKVIPHDKVYIRNVIISGNSRTVDVRIRRELYIT 368
Qy 248 PGTPYDLDLLLFOQALEQNGHYSGASVOADFRLDRGVRPVKSVTEVKRHKLETGIRL 307
Dy 369 EGNLYARTDLESKNALKRTSYFDDVNIKE--EKVDHIDHLLVDVKEASTGAISGIGY 426
Qy 308 DSEYGLGGKIADYDYNLFNKGYIGSVWDMDKYETTLAAGIS--OPRNYRGNTWTSNVS 365
Dy 427 GSSDILLNASLSDTNIFSGIKSSV--SVDSDDTLSGRISLVNPRVLSQYSLGGTLY 484
Qy 366 NRSTTONLEKRAFSGGIW--YVRDRAGIDARLGAFF-----LAERKIPGSDI----- 411
Dy 485 SNDE-----WDNSEKNYGFDTITIGROFARYNVNLSUTYNLESDIYHLSPTL 532
Qy 412 ----DLGNHATMLTASWKRQLLNVLHPNGHYLDGKIGTTLGTLSLTALIRTSARA 466
Dy 533 LRTGYELGKSIKSSITPAITENDTDYVLPKSG-----IIASTSLEYAGLGG 579
Qy 467 GYFPTPENKKLTF-----IIRCOAGYTVARDNADVPGLMFRSGGASSVGRYE 515
Dy 580 DQEFISSSKFNFYQGLQYIGYDLYIRYKASFYKFWDEGLYPLNQRIYLGGRINSRNGFE 639
Qy 516 LDSIGLAGPNSVLPERALLVGSLEYQLPFTR--TLSGAVFHDMDGAAA-----NFRMKL 569
Dy 640 SRTVSPKNQWDEIGGTIAFANSVELSPFLIDRIKRGSVFFDYGMIGRKNLDEIKRM-- 697
Qy 570 KHSGSLGVWFPSLAPPSFDIAYGHSKK---IRWHISLGRF 609
Dy 698 --STGIEIWTPIGQLVFAKPLNDKKGDDTNSFEFNLGRF 739

RESULT 35
P73472 PRELIMINARY; PRT: 861 AA.
AC P73472;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Chloroplast import-associated channel IAP75.
GN IAP75 OR SLR1227.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OC NCBI_TaxID=1148;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiyama M., Sugiyama S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR ENBL: D90906; BAAL7512.1;
DR InterPro: IPR000184; Bac_surfag_d15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
DR TIGRFAMs: TIGR00992; 3a0901s03IAP75; 1.
```

```
KW Complete proteome.
SQ SEQUENCE 861 AA; 92307 MW; 1C5DFFD394DA5B44 CRC64;

Query Match
Best Local Similarity 6.4%; Score 202.5; DB 16; Length 861;
Matches 131; Conservative 77; Mismatches 254; Indels 179; Gaps 23;

QY 76 TQOQEEVLDEQGTFLAEAPDNVKTMLRSKGYFSSKYSLETKDGAVYVHTIPGTRKIA 135
DB 293 TQIQEDV-----NAIVATGYFSN-VRVAPSDTPILGVRVTFEVQANPV 333

QY 136 NVGV-----AILGDILSDGNLAAYRNALENMQPVGSDPDQSDWNSKTSV 182
DB 334 FTGLNIRTPETAEGKERILUQEVVDTEFGQYKILNRLRELQEGIKTINEWYSN-----388

QY 183 LGAVTRGYPPLAK-LGNRAAVNPDTATVDLNVVDSGRPTAFGFEITGTQRYPEGLVS 241
DB 389 -----QGYDLAQVVGSPQVAGCGQVTLVIAEGIVENIQ-VREFDSE-----DEPQV 433

QY 242 GLARFQGTPTVDLILLDFQQALEQNGHYSGASVQADFRLQG-----284
DB 434 GRTR-----DFITTRMRLKPGDVFNNRRAQTDLQRYVSLGLEFEDVRLSENPQSDP 484

QY 285 DRVPKVSYTEVRKHLETKIRLDSYGLGGKIAYDYNNLKNKGYIGSVVNDMDKYETTL 344
DB 485 TEVINVDVVEGNTSGIAAGGGISSSSGLFTGISQERNLGNNOTICGEAQVQORELLF 544

QY 345 AAGISQPRNRYGNWTSNVSYNRSTTONLEKR-----AFSGGIWYVRDRAGIDA-----393
DB 545 DVSTDP-----WIGGDPFTSYTANLFRRTISLVPQDAGSSIRTFNGFDSPRVVRT 597

QY 394 -----RLCAEFLAEGRKI-----PGSDIDL 413
DB 598 GLGLTFPRITADDFAPPDWRLSAGFGYQNVRIENNAGALSPFSNPLNGFSQPLSFSY 657

QY 414 GNSHATMLTASWKQLNNVLPENSHYLDGIGTTLGTFLSSTALITSARACY-FFTP 472
DB 658 GVDELFTLSFGASQDNRNNALQPTSGSLV--RFGAEQIPVGTGNIMTRLGSYSYIIP 715

QY 473 -----ENKLGTFIIRGQAGYVARDNADVPGLMFRSGGASSVRGYELDSIG 520
DB 716 VNWLDTLGFGLVESTQPTVAFNVQAGTVL---GDLPPYEAFLGGSNSVRGYQEGELG 771

QY 521 LAGPNSGLVPERALLVGSLEYQLPFTRTLSGAVFHDMDG-----AANFRM-----KL 569
DB 772 -----NG-----RSFFQATAYRFPITAAVGGALFVDYNSLGSOGAVPGFPAIVRGLPGS 822

QY 570 KHGSLGVRFSPFLAPSFSDIAY-GHSDKKIRWHISLGLTRF 609
DB 823 GVGYGLGVRIQSPVGPTRIDLGFTGEGESRI--NFGICEKF 861

RESULT 36
Q9PKF3 ID Q9PKF3 PRELIMINARY; PRT; 792 AA.
AC Q9PKF3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Outer membrane protein, putative.
GN TC0512.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunnham R.C., Shen C., Gill S.R., Heideberg J.F., Bass S.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
```

```
"Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
pneumoniae AK39."; 28:1397-1406(2000).
RL Nucleic Acids Res.
DR EMBL; AE002320; AAF39354.1; -.
DR TIGR; TC0512; -.
DR InterPro; IPR000184; Bac_surfaG_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 792 AA; 88824 MW; 18387EC35B0F5666 CRC64;

Query Match
Best Local Similarity 6.3%; Score 199; DB 16; Length 792;
Matches 136; Conservative 76; Mismatches 232; Indels 174; Gaps 30;

QY 66 DMVEHLPLITIQOEEVLDEQGTFLAEAPDNVKTMLRSKGYFSSKYS---LTEKQG-- 120
DB 228 DMVDQDLFAIT-----NYFQNKGYADAKYTKVESTDAKGNL 263

QY 121 AYTWHITPGPTKIANY---GVAIGDILSDGNLAAYRNALENMQQVPGSD---FQDS 174
DB 264 SLVINVDKGPLYTLGHVHIEGFTALSKRLDKQLL-----VGNLSLYCPKV 310

QY 175 WENKTSVLGAVTARKGYPLAKLGNTRA---AVNPDTATVDLNVVDSGRPTAFGFEIT 230
DB 311 WAGAQ-KIRNAYARYGV---NTVDVSVFAVHPTLPVDYTVRVSEGAPYKIGLIKIK 364

QY 231 CTORYPQOIVSGLARFQPGTPYDLDLLDFQQALEQNGHYSGAS---VOADFRLQDDR- 286
DB 365 GNTTKHDVILHETSLFPGDTEFLKLEDTETRLRNTGYFKSVSYVTYRSQDPLDSNNL 424

QY 287 ---VPVKVSYTEV-----KRHKLETG---IRLDSE 310
DB 425 YRDVPFIEVKETETCNLGLFLGFSIDHLCGAEISHSNPDLCARHLFKGFKSLRGGE 484

QY 311 Y-----GLGKIAYDYNNLKNKGYIGSVVW---DMKYETTLAGISQPRNRYGNWTS 361
DB 485 YLFKANLGDKVIT-DYTVKTKPHFLNTPMILGVLDK---SINKLSK-----DYSVD 534

QY 362 NVSNRSTQTKLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATML 421
DB 535 TVGGINSTIYILNDK-LKYGMYV---RGSQTSLSLRKKTENPNKPGDLD---SNKGFV 586

QY 422 TASWKROLLNNVLHPEN-----CHYLDGKIGTTLGTFLSSTALITSARACYFPT 472
DB 587 SAAGLNVLYSIDNPRKPTMGIRSSLNFELSLGGLGTYOPTKLT-----SGSIYRL 637

QY 473 ENKLGTFIIRGQA-----GYTVARDNADVPGLMFRSGGASSVRGYELDSIG----LA 522
DB 638 LTKK-GVLKIRGEAKFKPFGTTAQ---GIPVSEFFLGGESTVRGYKPPFLIGPKESPT 693

QY 523 GPNGSVLPERALLVGSLEYQLPFTRTLSGAVFHDMDGDAANFRKMKLKH---GSLGV- 578
DB 694 EPQGL--SSLLTTEEFQYPLISQPSINAPVFLDSGFIGTEETIIRLKNLACSSAGFLRF 751

QY 579 -----WFSPLAP 585
DB 752 DMNNVPITMLGWGWPFRP 769

RESULT 37
Q9PEI2 ID Q9PEI2 PRELIMINARY; PRT; 784 AA.
AC Q9PEI2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Outer membrane antigen.
GN XF1046.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group.
OX NCBI_TaxID=2371;
RN [1]
```


Qy	449	TLGTF--LSSTALIRTSARAGYFTFPENKLGFTFIIRGOAGYTVARONADVP	SL-----M	503
Db	612	QISKYWIPIALVLNLRLEVGY-----GGDYKSHTRILPDGTV	-----TASGLPF	661
Qy	503	FRSGGASSVRGYELDISGLAGP-----NGSVLPERALLVGSLE-----	YQLPFTRLS	550
Db	662	FYAGGTNSVGRFRDNTLGRPSEVTALYNQOPLGGSEKTVGSTEMYFPKLF	DSARISA	721
Qy	551	GAVFHMDGAAANFRKMKLKHGSLGVNRPSP	LAAP-----FSPDIAYHSGKKIKHHSISLG	606
Db	722	FLDGFNGVFNKNFANELRASSCVALLNAPIGPISISYAFPIKKNENDEIR	LOFTFG	781
Qy	607	TRF	609	
Db	782	GQF	784	
RESULT 38				
Q8YHHO		PRELIMINARY;	PRT;	781 AA.
AC	Q8YHHO;			
DT	01-MAR-2002 (TRENBLrel. 20, Created)			
DT	01-MAR-2002 (TRENBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)			
DE	Outer membrane protein.			
GN	BMEI0830.			
OS	Brucella melitensis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Brucellaceae; Brucella.			
OX	NCBI_TaxID=29459;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=16M / ATCC 23456 / BIOTYPE 1;			
RX	MEDLINE=20020109; PubMed=11756688;			
RA	DeIvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.			
RA	Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,			
RA	Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,			
RA	Salkov E., Elzer P.H., Hagius N., O'Callaghan D., Letesson J.-J.,			
RA	Haselkorn R., Kyrides N., Overbeek R.;			
RT	"The genome sequence of the facultative intracellular pathogen			
RT	Brucella melitensis."			
FL	Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).			
DR	EMBL, AB009524; AAL52011.1;			
DR	InterPro: IPR000184; Bac.surfAg_D15.			
PF	Pfam: PF01103; Bac_surface_Ag; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 781 AA; 85918 MW; E3C7385CCB11DB3A CRC64;			
Query Match				
Best Local Similarity 19.5%; Pred. No. 0.00064;				
Matches 141; Conservative 115; Mismatches 284; Indels 183; Gaps				
Qy	26	LSENKAAGPALF-----KKNSPD-TESVKLKPKEPPVIDQDSEIKDWEEHLPIITQQQE	80	
Db	103	VKERSVYNNVLFGQNHKIKDPLARAYQLKPRAPFDNATMEAD-----	-----K	146
Qy	81	EVLDDKEQTGFLAEAPDNVKTMLRSKGYFSKSVLSEKDGATVHTPGPRTKIANYGVA	140	
Db	147	EAIKAAYSHGSRDATVNATVDLGQG-----RVNV-----VYEINSGSRTKIANIEFV	195	
Qy	141	-----ILGDILS---DGNLAEEYRNALENWQPVGSDFDQDQSWENSKTSVLGAVTRK	189	
Db	196	GNQAFSGRRIRDVITKRSNPLSLMTNDV-----YDEGRLOADEETLRRFYNNR	245	
Qy	190	GYPLAKLGNTRAAVNPDPATVDLNVVYDSDGRPIAFGDFEITGT-ORYPEQIVSGLARFOP	248	
Db	246	GYADPRVLSNAVLDPSTNEVITITVDEGPRYTFGDSVESTVDGSDTQALDLVKTRT	305	
Qy	249	GTPYDL-----DLLLDFQALQONCHYSGASVQADFDR-LQGRVPVKVSYTEVHRKLE-	302	
Db	306	GKPYSAKETDSVLSVTESSVAGSG-YAFAKVEPRGRDNFNENHTISVYVSDQGPRVYIOR	364	
Qy	303	---TGIRLDSYGLGCKTAYDYNNLFNKNYIGSV---VVMDDKYET---TLAAGISQPR-	352	

```
Db 365 IEIRGNDKTRDYVIRREFDLNEGDAPFNOVMQRAKRRLVLDFFQTVNISTAPG-SEPDQ 423
Qy 353 -----NYRGNWTSNVYNRSTTONLEKRAFSGGIVWVRAG----- 390
Db 424 VILVVDVVEKSTGEFSGIGGTYTGGESPAQVEAAIERNFLGRGVIRISAGQDDMR 483
Qy 391 -----IDARLGAEFLAEGRK-----IPGSDI----- 411
Db 484 NYGLSFTEPFYGLVRLSAGFDVRRSYRVNDYDVEQTGRTRECLPITDNFSAGIAYSL 543
Qy 412 -----DLGNSIATML-----TASWKR-----OLLNNVLHPENGHYLDKIG 447
Db 544 VQEKYDLFRGDAENYAPALLEAENSPWLRSSVSYSITYSIDDIKNPHDGLY--GKFI 601
Qy 448 TTLTFLSSTALIRTSARAGYFTPENKLGTFIIRGOAGYTVARNADVPVSGLMFRSGG 507
Db 602 QEPAGLGDAKYVKTFTKGNYYOTLSQADIVGLLGVGAGYHFEFGDDCYRIFDLFKN-S 660
Qy 508 ASSVRGVELDSIG--LAGPNS--VLPERALLVGSLEYQ-----LPFTRLSGAVF----- 554
Db 661 SDIIRGFKFGICGPYQDAKNGKRYMGGTTYESGTAEQVFPMPVLPSGLVGRGAFFADAA 720
Qy 555 -----HDMGDAANFKRMKLKHGSLGVRWFSPAPFSDIAY-----GHSDDKIRWHISLG 606
Db 721 TLYCNDTPDISGDDK--KLRSVGVSLMWASPFGLREDYAFPAKADTKVQNFNGVS 778
Qy 607 TRF 609
Db 779 TKF 781

RESULT 39
Q93PM2 PRELIMINARY; PRT; 793 AA.
AC Q93PM2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Integral outer membrane protein.
DE D15.
GN Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OC NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21295095; PubMed=11401984;
RA Thomas K.L., LeDuc I., Olsen B., Thomas C.E., Cameron D.W., Elkins C.;
RT "Cloning, Overexpression, Purification, and Immunobiology of an 85-
RT Kilodalton Outer Membrane Protein from Haemophilus ducreyi.";
RL Infect. Immun. 69:4438-4446(2001).
DR EMBL; AF329831; AAK70345.1;
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
SQ SEQUENCE 793 AA; 88661 MW; E4E9DB62A8F9903F CRC64;

Query Match 6.1%; Score 193.5; DB 2; Length 793;
Best Local Similarity 19.2%; Pred. No. 0.00076;
Matches 150; Conservative 104; Mismatches 278; Indels 249; Gaps 30;

Qy 27 SENKAAGFALKKNSPDSTESVCLKPFPV-----IDTQDSEKQWVEEHLPLITQQQ 79
Db 64 SONR-----FENVSAREQTLVTKVAERPLINLTTKGNAIPKNALEONKANLIVA 117
Qy 80 BEVLDEKQGTGLAEAPADNVKTMRL-----SKGVFSKSVLSTEDKDAYT----- 123
Db 118 GEVVDKAKLEAFKQALVDHYHTMGVQADIQTITPPNNGNSINVELNITEGEIAYVKIN 177
Qy 124 -----VHITGCP-----RTKIANGVAILGDILSDGNLA----- 152
Db 178 FEGNNAFSYDELKELKPNAPMWNIFESKFOQOENKDIILRDPYMDHGYAKFTLK 237
```

```
Qy 153 -----EYVRNALE--NWOOPVGSDF-----DODSWENSKTSVL-----GAVTRK-- 189
Db 238 DTQVKFNEKTEVDLYTKINEGSOYNISEMRIIGDTQKLDNELNQLLTHFKAGOLFPRKE 297
Qy 190 -----GYPLAKGNTRAAVNPOTATVDLNVVVDGSRPIAFGDFEITQTR 234
Db 298 LSIIEEQIKOILGDRGYGSAKV--DLYPKFNEEDHTVQINFIVDAGRRIYVRKIRFEGNDV 356
Qy 235 YPEOIVSGLARFOPGTPYDLDDLDFQOALEONGHYSGASVQADFDRLQGRVVPKVSVT 294
Db 357 TADSTLREMRQEGAWLSTSAVSLAKSRLERTGFYEVMSMPTVKNTDQDDVIYKIK 416
Qy 295 EVKHKHLETKIRLSDSEYGLGKRAYDYNNLPMKGYIGSVVMDMKYETTLAAGISQPR-- 352
Db 417 ERNTGSINFGV--GYGSGGSLSYN-----AGITODNFL 447
Qy 353 -----NYRGNWTSNVYNRSTTO--NLEKRAFSGGIWYVRORAGIDARLGAFLAEG 403
Db 448 GMGSSILGNSRNTDSTNVNLSYTEPYFTKDGVSIGGNIFY-----EDYDNRKASAAAYK 503
Qy 404 RKIPGSDIDLG--NSHATMLTASWKRQLNNV-----LHPENGHYLD 443
Db 504 RKTYGASGTGLGFPVDENNYSYLGGLGYTHDKLRNVEREYTRKRYVNSMKFPIPNQSHY-- 561
Qy 444 GKICITLGLTSLALIRTSARAGYFT-----PENK--KLGT----- 479
Db 562 DRIOS--ADFOLSEGNYNLNLRGYEPTAGSSANISGKLTLPDSDNKYYQVGTNFSGYIP 619
Qy 480 -----FIIRGOAGYTVARNADVPVSGLMFRSGGSSVRGYELDSI-----GL 521
Db 620 LNSEHKWVIATKGLAYTNSFGGKEVPFYQLYSAGSMGLRGFAGGSGIPKAIYYREDGF 679
Qy 522 AGPNSVLPERALLVGSLEYQLP-----PTRLTSGAVF----- 554
Db 680 KAPSDQVIGGNANVNASLELIIPAPFISDKYQHNVRTSVEYDAATVWNTKKWOSKADYPN 739
Qy 555 -HDMGDAANFKRMKLKHGSLGVRWFSPAPFSDIA-----YGHSDKKTRWHISLQTR 608
Db 740 LPDFGD----YKRVRA--SAGIALQWQSPIGLSPFYAKPIKKA-GDEIQFOFTVGST 792
Qy 609 F 609
Db 793 F 793

RESULT 40
Q92Q48 PRELIMINARY; PRT; 776 AA.
AC Q92Q48;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative outer membrane transmembrane protein.
GN OMP OR R01502 OR SMC02094.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OC NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021
RX MEDLINE=21336507; PubMed=11481430;
RA Capela D., Barloy-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallbert P.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9677-9682(2001).
DR EMBL; AL591787; CAC46081.1;
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
```

Search completed: November 9, 2002, 01:20:14
Job time : 39 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:14:29 ; Search time 17 seconds

(without alignments)

1485.828 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MMKPTALLPALFFPPHAY.....IAYGHSDKKIRWHISLCTRF 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	379.5	12.0	577	1 YTFM_ECOLI	P39320 escherichia
2	374	11.8	578	1 YTFM_HAEIN	P44038 haemophilus
3	217	6.8	810	1 UP05_ECOLI	P39170 escherichia
4	213.5	6.7	793	1 D153_HAEIN	O32629 haemophilus
5	209.5	6.6	795	1 D152_HAEIN	P44935 haemophilus
6	209.5	6.6	797	1 D151_HAEIN	P46024 haemophilus
7	138	4.3	835	1 PASD_ECOLI	P46000 escherichia
8	133	4.2	779	1 K6PF_HUMAN	P08237 homo sapien
9	131	4.1	781	1 K6PF_CANFA	P52784 canis famill
10	128	4.0	779	1 K6PF_RABIT	P00511 oryctolagus
11	124	3.9	881	1 YEBT_HAEIN	P44288 haemophilus
12	121.5	3.8	617	1 Y237_BUCAI	P57331 buchnera ap
13	120	3.8	779	1 K6PF_RAT	P47858 rattus norv
14	118.5	3.7	815	1 FPVA_PSEAE	P48632 pseudomonas
15	118.5	3.7	815	1 K6PF_MOUSE	P47857 mus musculu
16	116.5	3.7	1300	1 120K_RICRI	P14914 rickettsia
17	116.5	3.7	1654	1 OMPE_RICRI	O53047 r outer mem
18	116	3.7	922	1 PMPI_CHLPN	Q95965 chlamydia p
19	114.5	3.6	3035	1 APLP_MANSE	Q23490 manduca sex
20	113.5	3.6	1018	1 VGNM_BPMV	P23009 bean-pod mo
21	113.5	3.6	1577	1 HLYA_PROMI	P16466 proteus mir
22	113	3.6	838	1 YRAJ_ECOLI	P42915 escherichia
23	111.5	3.5	595	1 VPI_BPCHP	P19192 bacterioph
24	111	3.5	574	1 TACY_STRCB	O53957 streptococ
25	111	3.5	976	1 FTBP_ADEB3	O03553 bovine aden
26	110.5	3.5	691	1 YHUG_ECOLI	P37645 escherichia
27	110	3.5	828	1 MRKC_KLEPN	P21647 klebsiella
28	110	3.5	1423	1 FRUA_STRMU	O03174 streptococ
29	109	3.4	571	1 TACY_STRPY	P21131 streptococ
30	109	3.4	584	1 FHAC_BORPE	P35077 bordetella
31	109	3.4	1076	1 CARB_VIBCH	O9kph9 vibrio chol
32	108.5	3.4	841	1 MYFC_YEREN	P33408 yersinia en
33	108.5	3.4	1375	1 GTFC_STRMU	P13470 streptococ

ALIGNMENTS

RESULT 1

ID	YTFM_ECOLI	STANDARD;	PRT;	577 AA.
AC	P39320;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein ytfM precursor.			
GN	YTFM OR B4220 OR Z5831 OR EC55198.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=95334362; PubMed=7610040;			
RA	Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,			
RA	Blattner F.R.;			
RT	*Analysis of the Escherichia coli genome VI: DNA sequence of the			
RT	region from 92.8 through 100 minutes.*;			
RL	Nucleic Acids Res. 23:2105-2119(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RT	*Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;			
RT	Nature 409:529-533(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / RIMD 050952;			
RX	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,			
RA	Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	*Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT	O157:H7 and genomic comparison with a laboratory strain K-12.*;			
RT	DNA Res. 8:11-22(2001).			
CC	!- SIMILARITY: STRONG. TO H INFLUENZAЕ H10698.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).			
CC	-----			

34	108	3.4	517	1 HSP2_MOUSE	P38533 mus musculu
35	108	3.4	574	1 FLA3_CAMJE	Q46113 campylobact
36	107.5	3.4	479	1 PRTC_ERWCH	P16317 erwinia chr
37	107	3.4	587	1 PICP_PSESR	P42790 pseudomonas
38	107	3.4	791	1 K6PP_RABIT	P47859 oryctolagus
39	107	3.4	859	1 AFAC_ECOLI	P53517 escherichia
40	106.5	3.4	356	1 E2BI_PYRAB	O9uz16 pyrococcus
41	106.5	3.4	1320	1 PUTA_ECOLI	P09546 escherichia
42	106.5	3.4	1655	1 OMPE_RICCN	O9kka3 r outer mem
43	106	3.3	3034	1 CLRL_MOUSE	O35161 mus musculu
44	105.5	3.3	812	1 FAED_ECOLI	P06970 escherichia
45	105	3.3	574	1 FLB3_CAMJE	Q46114 campylobact

ENBL; U14003; AAA971116.1; -
ENBL; A600493; AAC71717.1; -
ENBL; A6005654; AAG59418.1; -
ENBL; A6002568; BAB38621.1; -
Ecogene; EG12513; YTFM.
InterPro; IPR000184; Bac_surfa_d15.
Pfam; PF01103; Bac_surface_Ag; 1.
Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 577
FT SIGNAL 1 21
FT CHAIN 22 577
SEQUENCE 577 AA: 64796 MW; 253DID5BEA74D25 CRC64;
Query Match 12.0%; Score 379.5; DB 1; Length 577;
Best Local Similarity 24.7%; Pred. No. 3.3e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;
QY 74 LIHQOEEVLKDEQTEFLAEAPDNVKTMLRSKGYFSSKVSF---TEKDAYVY---HIT 127
Db 43 LSTIEDEVTDPDR---FRARVDDAIRESKALGYQPTIEFDLRPPPKQRQLIAKVT 99
QY 128 PGPRTKIANGVAILGDILSDGNLAIEYRNALENNQOQVGSDFQDQSWENSKTSVLGAVT 187
Db 100 GGPVVLIGGTGVVLRGARTD---KDYLFK---LLDTRPAIGVNLNOGDYENFKSLTSLAL 154
QY 188 RKGY-----PLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGRTORYEQIVSG 242
Db 155 RKGYPDFSEFTKAQLG---JALGLHAKAFWDID---YNSGRYRFGHVTFEGSO-IRDEYLN 208
QY 243 LARFPQGTPTDLDLLDFQOALEQNGHYSGASVQADFDRLQGDRTV-PVKVSVTEVKRHL 301
Db 209 LVPFKGDEYKDLAELNRLRSATGWSNVVAPQFOKARETKVPLTLGVVSPRTENTI 268
QY 302 ETGIRLDSVGLGGKTAIDYDYNLFNKGYIGSVVMDKAYETTLAAGISOPRNYKGNWTS 361
Db 269 ETGCVGSTVDGPRVATW-----KKPWNYSY-----CHSLTSTTSIAPEQ-----TL 311
QY 362 NVSYNSTRSTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATML 421
Db 312 DFSYKMPLLKN-----PLEQYVLVQG-GFKRTRDLMDTESDSTTL 349
QY 422 TAS-----WKROLNNLHPENGVLDGKIGTIGTLGTFSLSTALRTSARACYF----- 469
Db 350 VASRYWDLSSGWORAI---NLRWSLDHFOTGEITWTMLFPGVWISRTSRSGGLMPTWG 406
QY 470 -----FTPENKLGTF-----FIIRGAGYTVARDNADVPSSL 501
Db 407 DSRVSYDYSNTAGSDVDVSVFOAQNWIRTLVDRHFRVTRGLGMIETGDFKVPDDL 466
QY 502 MFRSGASSVRGVELDSDIGLAPNGSVLPERALLVGSLEYQLPTFRTLSGAVFHDGMDAA 561
Db 467 RPPAGGDSIRGYKYKSIAPKANGDLGASKLITCSLEYQYNYTKWNGAVFVDSGEAV 526
QY 562 ANFKRMKLKHSGLGVRFSPSLAPSFDIACHSDKK---IRWHISLG 606
Db 527 SDIRRSDFKTGTGVGVWESVPGPKLDFAVPVADKDEHGLQFYIGL 574
RESULT 2
YTFM_HAEIN
ID YTFM_HAEIN STANDARD; PRT: 578 AA.
AC P44038
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein HI0698 precursor.
GN HI0698
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxId=727;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;

```

Db 376 QADITDKTLLLYPTVGVTRTLRGGSPATWG-----DVQKI---TFDLS----- 416
QY 426 KQOLLNVLHPENGHYLDGKIGTTLTGLTFLSSALIRTSARAGYFTT-PENKKLGTFTIRG 484
Db 417 KR-----TWLSESSFIKVQASSAWRTYAENHRV---VARA 449
QY 485 QAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAPNGSVLPERRALLVGSLEYOLP 544
Db 450 EIGYLHTGIEKIPPTLRFPAGGDRSVRGYKKIAPKRNKGLVGGSRLLTTSLEYOYQ 509
QY 545 FRTLGSVAFHMDGAAANRKRKLKNGSLGVWRFSPLAFSFDIAYGHSD-----KKIR 600
Db 510 VYPNMAATFADSLAANDYAKELRGYTGVRWASPVGAIKFDIATPIRDKONSKNIQ 569
QY 601 WHISLGT 607
Db 570 FVIGLGT 576

RESULT 3
UP05_ECOLI
ID UP05_ECOLI STANDARD; PRT: 810 AA.
AC P39170; P39181; P77465;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page spots M62/M63/O3/09/T35 precursor.
GN YAE1 OR B0177 OR Z0188 OR EC50179.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).

```

```

RN SEQUENCE OF 21-32 AND 351-362.
RC STRAIN=K12 / EMC2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12."
RL Electrophoresis 18:1259-1313(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000127; AAC73288.1;
DR EMBL; U70214; AAB08606.1;
DR EMBL; AE005193; AAG54479.1;
DR EMBL; AP002550; BAB33602.1;
DR SWISS-2DPAGE; P39170; COLI.
DR Ecogene; EG12676; yaeT.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 810
FT SEQUENCE 810 AA; 90552 MW; DDC64C6D341664EB CRC64;
SQ
Query Match 6.88; Score 217; DB 1; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.6e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;
QY 97 DNKVTMLRSKGYF-----SSKVSILT-EKDGAY-TVHITPGRPKTKIANGVAILGSDG 149
Db 224 ETLLRSYLRDGRYARFNIDSTQVSLTPDKKGIYVTNITGGDKYKLSGVEVS-----G 275
QY 150 NLAEYRNALENWQOPVGSDFDQDSWENSTSVLGAIVTKGYPLAKLGNTRAAVNDPAT 209
Db 276 NLAGHSABEITLTKEBGLYNGTKVTKMEDDITKLLGRGYAYPRV-QSMPEINDAKT 334
QY 210 VDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFPQPTPYDLDLLDFOALEONGH 269
Db 335 VKLRVNDVAGNFRYVRKIRFEGNDTSKDAVLREMRQMEGANLGSGLVDQGERLNRLGF 394
QY 270 YGASVQADFDRLQG--DRVPKVSIVTEVRHKLETGIRLDSEYGLGGKIADYINLENK 327
Db 395 FE--TVDTDTQRPVGSQDVVDVYKVKERNVTGSFNGIGYGTESGVSFOAGVQDNWLG 452
QY 328 GYIGSVWMDMKYETTLAAGISQPRNYRGNYWT-----YFTVDGVSGLGRLFYNDFOADDLSDVTN 362
Db 453 GYAVGINTGKNDYQTYAELSVTNP-----YFTVDGVSGLGRLFYNDFOADDLSDVTN 506
QY 363 VSYNRSSTONL---EKAPFSGGIWYVDR-AGIDARLCA-EFLAEGRKTPGSDIDLGNH 417
Db 507 KSYGTDVTLGTPPINEYSLRAGLGYVHNSUNMQPQVAMWRYLYDSNGEHP-STSDQDN 565
QY 418 AT---MLTASMKRQLLNVLHPENGH--YLDGKI-----GTTLTGTLSTALIRT 462
Db 566 KTDDTFNYGVNTYNNKLDRGYFPTDGRSVNLTKGVTIPGSDNEYKVTLOT----- 615
QY 463 SARAGYFTTPENKKLGTFTIRGQA--GYTVARNADVPGLMFRSGGASSVRGYELDSIG 520
Db 616 ---ATYVPIDDDH---WVWLGTRWGYDGLGKEMPFYENFYAGGSSTVYRGFSNTI- 668
QY 521 LAGPNSGLVLP-----ALLVGSLEYQLP----- 544;
Db 669 --GPRAVYFPHQASNYDPDYECATQDGAKDLCKSDDAVGGNMAVASLEITITPTPTS 726

```

```

Qy 545 --FRTSLGAVFHDMDGA-AAANFKRMK-----LKHGSGLCGRWFSP LAPSEF 589
Db 727 DKYANSVRTSFFWDMCTVNDTWDSQSYGYDYSDPSNIRMSAGIALQWNSPLGPLVFS 786
Qy 590 IAVG----HSDKKIRWHISLG 606
Db 787 YAQPFKKYDGDKAEOQFNIG 807

RESULT 4
D152_HAEIN STANDARD; PRT; 793 AA.
AC Q32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RC MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U60834; AAB61977.1;
CC InterPro: IPR000184; Bac_surfAg_D15.
CC Pfam: PF01103; Bac_surface_Ag_1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A14 CRC64;

Query Match
Best Local Similarity 21.0%; Pred. No. 2.7e-06;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;

Qy 105 SKGYFSSKVSLT-----EKDGA-VTVHTPGPRTK-----IANVG--VAILGDILSDG 149
Db 228 NNGYAKAQITKTDVQINDEKTKVNVITDVNEGLQYDLSARIICNLGMSAELEPLLSAL 287
Qy 150 NLAEYVRNLENWQPVGSDFDSDSWNSKTSVLGAVTRKGYPLAKLGNRAAVNPD--- 206
Db 288 HLDNTPRS-----DIADVENAIKALIG---ERGY-----GNTVNSVPDFDD 327
Qy 207 -TATVDLNVVDSGPIAFGPEITGTQRYPEQIVGLARPGTQPYDLDLDDLLDFOQALE 265
Db 328 ANKTLAITFVVDAGRLTVRQLRFEGNTVVSADSTLQRMQEQECTWYNSQLVELGKIRLD 387
Qy 266 QNGHYSGASVQADFRLQG---DRVPVKVSVTEVKRKHLEGTIRLSEYGLGKGIADYDN 323
Db 388 RTGFFE--TVENRIDPINGSNDEVDDVYKVKERTGTSINGIGYGTSGISVQTSIKQDN 445
Qy 324 LFNKYGIGVWDMKXYETTLAAGISQPR-----NTRGNVWTSVNSYRNS--TTOMLEKR 376

```

```

Db 446 FLCTGAASVAGTCKNDYCTSVNLGYTEPYFTKDGSLGGNIFPENYDNSKSDTSSNYKRT 505
Qy 377 AFGGI---WYVDRAGIDARIG-----AEFLAE-----GRKIPGSDIDLG 414
Db 506 TYGSNVTLGFVPVNNNSYYVGLGTYTNKISNFALEYNRNLYIQSMKFGKNGIKTNDFD-- 563
Qy 415 NSHATMLTASWKRQLLNVLNHPENGYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPEN 474
Db 564 -----FSFGWYNSLNRGYFPTKG--VKASISLGGRTVTPGSDNKKYKLSADVGFPPLDR 615
Qy 475 KKLCTFIIRCOAGYTVARDNADVPGLMFRSGGASSVRGVELDSIGLAGPN----- 525
Db 616 DHRWVYSKASACAYANGFCNKLPLPYQTYTAGGICSLRGAYGSI---GPNAIYAEHNG 672
Qy 526 -----GSVLPERALLVGSLEYQLP-----FRTT-----L 549
Db 673 TENKISSDVIIGNAITASAEILVTPFPVDSKQNTVTSILFYDAASVWNTKKSKDKNGL 732
Qy 550 SGAVFHDMDGAAANFKRMKHLKHGSGLCGRWFSP LAPSEF 604
Db 733 ESKVLKDLDPYG---KSSKIRASTGCGFQWQSPICGLPVFSYAKPIKKYENDVYE-QQFS 788
Qy 605 LGTRF 609
Db 789 IGGSF 793

RESULT 5
D152_HAEIN STANDARD; PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
GN HI0917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ed / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Kerlavage A.R., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA McKenney K., Sutton G., Fitchugh W., Glodek A., Kelley J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32773; AAC22575.1;
CC TIGR; HI0917;
CC InterPro: IPR000184; Bac_surfAg_D15.
CC Pfam; PF01103; Bac_surface_Ag_1.
CC Antigen; Outer membrane; Signal; Complete proteome.

```



```
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 795 AA: 87478 MW: 85691FC22BED44 CRC64;

Query Match 6.6%; Score 209.5; DB 1; Length 795;
Best Local Similarity 20.0%; Pred. No. 4.9e-06;
Matches 130; Conservative 103; Mismatches 27; Indels 141; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOBEVLDKQGTGLAEAPDNVKTMLRSKGYFSSKVSUT--- 116
DB 184 ESVSSSTLQEQMELQPDSSWKLGNKFEQAQFEKQLQSDIRDYLLNNGYAKAQTITKTDVQL 243
QY 117 --EKDGA-YTVHITPGPRTK-----IANVG--VALGDLISDGNLAERYNALENWQOP 165
DB 244 NDEKTKVNWITDVNEGLODYLRGARIIGNLGGMSAELEPILLSALHLNDFRRS----- 296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKGNTRAAVNDP-----TATVDLNVVVDGRP 221
DB 297 -----DIADVENAIKAKLG--ERGVGSATVNSV-----PDFDANKTLATLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARFQGTPTDLDLLDFQQAQLEQNGHYSGASVQADFDR 281
DB 344 LTVRLQREFGNTVSADSTLRQEMRQOQGTWYNSQLVELGKIRLDRIGFFE--TVENRIDP 401
QY 282 LOG--DRVVPKVSVEVTKRHKLETGIRLSEYGLGKIAIDYNNLFNKGYIGSVVWMDK 339
DB 402 INGSNDEVDVYVKERNTGSIINFGYGTGESIGSYQASVKQDNFLGTGAAVSIAGTKND 461
QY 340 YETTLAAGISOPR-----NYRGNWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYTKDGVSLGGNVFENYDNSKSDTSSNYKRTYGSNVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKTPGSDIDLGNSHATMLTASWKROLL 430
DB 522 SYVYGLGHTYKNISNFALEYNNRLYIQSMKFKNGIKTNDP-----FSFGWYNSL 573
QY 431 NNVLHPENGHYLDGKIGTGLTGLTSFSLALRTSARAGYFTTPENKKGFIIRGQAGTV 490
DB 574 NRGYPPTKG--VKASIGGRVTIPGSDNKYKLSADVOGFPYPLDRDHLWWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPERA 531
DB 632 GFCNKRLLPFYTYTAGIGSLRGFAVGSIGSI---GPNAIYAEHNGNGTFKISSDVIGNA 688
QY 534 LLVGSLEYQLP-----FTRTISGAVFHDMDGAAANF----- 564
DB 689 ITTASAEILVPTPVSDKSONQVTRTSLF---VDAASVWTKWKSGLDNNVLKSLPDY 745
QY 565 -KRMLKHKHSGSLGRWPSPLAPFSFDIA-----YGHSDKKIRWHISLGTFR 609
DB 746 GKSSIRASTGVGFQWQSPICPLVFSYAKPIKYEYNDVE-QFQSIGGSF 795

RESULT 6
D151_HAEIN STANDARD: PRT: 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=Serotype B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack P.S., Loosmore S., Chong P., Thomas W.R.;
RT *The sequencing of the 80-kDa D15 protective surface antigen of
Haemophilus influenzae.*;
```

```
Db 689 NATATASAEIIVTFFVSDKSNVTRVSLFVDAASVWNTKWKSKDKNGSLDKVLKRLPDVG 748
QY 568 ---KLKHGSGGLGVNFWFSLAFSPEDIA-----YCHSOKKIKRWHSISLGTFR 609
Db 749 KSRIRASTGVGFQWQSPGIPLVFSYAKPIKKYENDDVE-QFOFSIGGSF 797

RESULT 7
FASD_ECOLI
ID FASD_ECOLI STANDARD; PRT: 835 AA.
AC F46060;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Outer membrane usher protein fasd precursor.
GN FASD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94148769; PubMed=7906265;
RA Schifferli D.M., Alrutz M.A.;
RT "permissive linker insertion sites in the outer membrane protein of
RT 987P fimbriae of Escherichia coli.";
RL J. bacteriol. 176:1099-1110(1994).
CC - FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE 987P
CC FIMBRIAE SUBUNITS ACROSS THE OUTER MEMBRANE.
CC - SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC - SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L22659; AAA21827.1; -.
DR EMBL: U50547; AAB02687.1; -.
DR TrimerPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 835 OUTER MEMBRANE USHER PROTEIN FASD.
FT DISULFID 810 834 POTENTIAL.
SQ SEQUENCE 835 AA; 92354 MW; 2PECE6E2305274E1 CRC64;

Query Match 4.38; Score 138; DB 1; Length 835;
Best Local Similarity 17.84; Pred. No. 0.21;
Matches 104; Conservative 87; Mismatches 201; Indels 192; Gaps 22;

QY 56 RIDTQDSEIKDVVEHLPLITQOQBEVLDRKQGTFLAEAPDNVKTMLRSKGVFSKVS 115
Db 259 KIYNDENLNDRLRSYTPV-----RGIASSQAV 288
QY 116 TEKDQATV---HITPGRTKIANGVNAIGDILSDGNLAERYRNALENQQPVSD--F 170
Db 289 TIQQGVVILQKNVPPGPF-----INDFSLSGYCDLYVNIKEADGSEHSF 335
QY 171 DQDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPTAFGDFEIT 230
Db 336 IQ-----PFSTLPEMKREGV-----SCVEISLGHYNS 363
QY 231 GTCRYPEQIVSLGARFQGTPTFDLLDLDFQQALEQN-----GHYSGASVQAFD 280
Db 364 GATQYNNSEPFILYASWSRGYRNGMTLYSETIQSRKGYOLLGVGSLGDFGAVSGDASL 423
```

```
QY 281 RL-----QGRVVPVKYVTEVKRHKLETGIRD--SEYGLGKIAIYDYNLNFKNKYGS 332
Db 424 RANKYDKIHUSGOSYGLKYS-----KNKVDGTGTTVTLATYRYSTKDYFSNDFVSKNDVQ 478
QY 333 VVWMDKYETTLAAGISQPRNYRG-----NYMTSN-VSYNRRSTONLEKRAFSGGI 382
Db 479 YVWD-NRLKNRITLSLQSDLDYGSLSLIASQQNYMTSDYVSRFSLS-----HSFGWMD 532
QY 383 WYVRDRAGIDARLGAFLAEGRKIPG--SDIDL-----GNSHATMLTASKROLLNNVLHP 436
Db 533 IFESTFSLQCKREGDNALNNKNNKVFYSSIPUSKLGKNKNESTYSTLSNVTNNQV-- 590
QY 437 ENGHYLDGKIGTTLGTLFSLTALRTSARAGYFFTPENKKGIT----FIIRQAGYTVAR 492
Db 591 RNFATLAGKVPQSMAYRFSFGWANTEQSS-----NKALSVMWDGDLDDGSLGYTSSG 643
QY 493 DN-----ADVPGLMFRSGGASSYRGYEL--- 516
Db 644 KNRIDYSLGSSAIIYPWRLAIGSDSVINGAAVVETFTSIGIKVGGGETSLLGTAIVTS 703
QY 517 -----DSIGLAGNGSVLPERALLYGSLLEYQL 543
Db 704 MQPYTENRIDLDTQIPDDLEFISNASKKIVPEKGAIV-PVKYNL 746

RESULT 8
K6PF_HUMAN
ID K6PF_HUMAN STANDARD; PRT: 779 AA.
AC P08237; Q16814; Q16815;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=92009225; PubMed=1833270;
RA Yamasaki T., Nakajima H., Kono N., Hotta K., Yamada K., Imai E.,
RA Kuwajima M., Noguchi T., Tanaka T., Tarui S.;
RT "Structure of the entire human muscle phosphofructokinase-encoding
RT gene: a two-promoter system.";
RL Gene 104:277-282(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=89306675; PubMed=2526045;
RA Sharma P.M., Reddy G.R., Vora S., Babior B.M., McLachlan A.;
RT "Cloning and expression of a human muscle phosphofructokinase cDNA.";
RL Gene 77:177-183(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=89030023; PubMed=2822475;
RA Nakajima H., Noguchi T., Yamasaki T., Kono N., Tanaka T., Tarui S.;
RT "Cloning of human muscle phosphofructokinase cDNA.";
RL FEBS Lett. 223:113-116(1987).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 271-680 FROM N.A. (ISOFORM 2).
RC TISSUE=Muscle;
RX MEDLINE=90264379; PubMed=2140567;
RA Sharma P.M., Reddy G.R., Babior B.M., McLachlan A.;
RT "Alternative splicing of the transcript encoding the human muscle
```

RT Isoenzyme of phosphofructokinase. ;
 J. Biol. Chem. 265:9006-9010(1990).
 RN [6]
 RP SEQUENCE OF 1-27 FROM N.A.
 RC TISSUE=Muscle;
 RA MEDLINE=89306652; PubMed=2526044;
 RX Valdez B.C., Chen Z., Sosa M.G., Younathan E.S., Chang S.H.;
 "Human 6-phosphofructo-1-kinase gene has an additional intron
 RT upstream of start codon. ;
 RL Gene 76:167-169(1989).
 RN [7]
 RP REVIEW ON GSD-VII VARIANTS.
 RX MEDLINE=96055509; PubMed=7550225;
 RA Raben N., Sherman J.B.;
 "Mutations in muscle phosphofructokinase gene. ;
 RL Hum. Mutat. 6:1-6(1995).
 RN [8]
 RP VARIANTS GSD-VII GLN-99; ASP-208 AND HIS-695.
 RX MEDLINE=95126102; PubMed=7825568;
 RA Raben N., Exelbert R., Spiegel R., Sherman J.B., Plotz P.,
 Heinisch J.;
 "Functional expression of human mutant phosphofructokinase in yeast:
 RT genetic defects in French Canadian and Swiss patients with
 RL phosphofructokinase deficiency. ;
 RL Am. J. Hum. Genet. 56:131-141(1995).
 CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate -> ADP + D-
 CC fructose 1,6-bisphosphate.
 CC -1- COFACTOR: MAGNESIUM.
 CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
 CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
 CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
 CC -1- SUBUNIT: Tetramer. MUSCLE IS M4, LIVER IS L4, AND RED CELL IS M3L,
 CC M2L2, OR ML3.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -1- DISEASE: DEFECTS IN PFKM ARE THE CAUSE OF GLYCOGEN STORAGE
 CC DISEASE VII (GSD-VII) (ALSO KNOWN AS TARUI'S DISEASE); A DISEASE
 CC CHARACTERIZED BY EXERCISE INTOLERANCE WITH ASSOCIATED NAUSEA AND
 CC VOMITING. SHORT BURSTS OF INTENSE ACTIVITY ARE PARTICULARLY
 CC DIFFICULT. SEVERE MUSCLE CRAMPS AND MYOGLOBINURIA DEVELOP AFTER
 CC VIGOROUS EXERCISE. MOST PATIENTS OBTAIN A "SECOND WIND" WHEN THE
 CC ONSET OF EXERCISE IS FOLLOWED BY A BRIEF REST PERIOD. IN TIME
 CC PATIENTS ADJUST THEIR ACTIVITY LEVEL AND ARE WELL COMPENSATED.
 CC -1- MISCELLANEOUS: IN HUMAN PFK EXISTS AS A SYSTEM OF 3 TYPES OF
 CC SUBUNITS, PFKM (MUSCLE), PFKL (LIVER) AND PFKP (PLATELET)
 CC ISOENZYMES.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M59735; AA82938.1; JOINED.
 DR EMBL; M59736; AA82938.1; JOINED.
 DR EMBL; M59737; AA82938.1; JOINED.
 DR EMBL; M59738; AA82938.1; JOINED.
 DR EMBL; M59739; AA82938.1; JOINED.
 DR EMBL; M59740; AA82938.1; JOINED.
 DR EMBL; M26066; AAA60068.1; -.
 DR EMBL; Y00698; CAA68692.1; -.
 DR EMBL; BC000534; AAH00534.1; -.
 DR EMBL; J05533; AA79220.1; -.
 DR EMBL; M24225; AAA36436.1; -.
 DR PIR; PH0223; PH0223.
 DR PIR; S00158; S00158.
 DR PIR; J50205; J50205.
 DR HSP; P00512; 3PFK.
 DR HSC-2DPAGE; P08237; HUMAN.
 DR Genew; HGNC:9877; PFKM.
 DR MIM; 232800; -.
 DR MIM; 171850; -.
 DR InterPro; IPR000023; Ppfrckinase.
 DR Pfam; PF00365; PFK; 2.
 DR PRINTS; PR00476; Ppfrckinase.
 DR ProDom; P000707; Ppfrckinase; 2.
 DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
 KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
 KW Phosphorylation; Magnesium; Multigene family; Alternative splicing;
 KW Disease mutation; Glycogen storage disease.
 FT INIT_MET 0
 FT MOD_RES 774 774 PHOSPHORYLATION (BY SIMILARITY).
 FT VARSPIC 281 311 MISSING (IN ISOFORM 2).
 FT VARIANT 38 38 R -> L (IN GSD-VII; ASHKENAZI).
 FT VARIANT 38 38 R -> P (IN GSD-VII; ITALIAN).
 FT VARIANT 38 38 R -> Q (IN GSD-VII; SWISS).
 FT VARIANT 99 99 /FTID=VAR_006064.
 FT VARIANT 208 208 G -> D (IN GSD-VII; FRENCH CANADIAN).
 FT VARIANT 542 542 /FTID=VAR_006066.
 FT VARIANT 685 685 D -> A (IN GSD-VII; ITALIAN).
 FT VARIANT 685 685 W -> C (IN GSD-VII; JAPANESE).
 FT VARIANT 695 695 /FTID=VAR_006068.
 FT VARIANT 695 695 R -> H (IN GSD-VII; SWISS).
 FT /FTID=VAR_006069.
 SQ SEQUENCE 779 AA; 85051 MW; 4C9F3847A7A5A5750 CRC64;
 Query Match 4.2%; Score 133; DB 1; Length 779;
 Best Local Similarity 20.6%; Pred. No. 0.41;
 Matches 132; Conservative 71; Mismatches 241; Indels 196; Gaps 31;
 QY 81 EVLDKEQTGFLAEAPDINVKLSKGYFSKSVLTKEDGAYTVHTPGPRKIANVGYA 140
 DB 186 EIVDAITTAQSHQRTVLEVNGRHCYGLALVLSLSCGADWVFIPECPDP----- 235
 QY 141 ILGDILSDGNLAETTYRNALENQQPVGSDFDQSDWENSKTSVL-----GAVTKGYPLAKL 196
 DB 236 -----DDWEHLCHRLSETRTRGSRNLIIIVAEAGDKNGKPTSE 276
 QY 197 GNTRAAVNP---DTATVDLNVVDSGRPIAGDFEITGQRYPEQIVSGIARPOGTP-- 251
 DB 277 DIKNLVYKRLGYDTRVTVLGHVQRGGTFSAPD--RIILGS-RMGVEAVMALLEGTPDTAC 333
 QY 252 -----YDLDLIDL----FOALAEQNGHYSGASVQADFRLQGDVRV 287
 DB 334 VVSLSGNAVRLPLMECEVQVTKDYTKAMDEKKFDEALKLGR-SPMNNHEVYKLLAHVRP 392
 QY 288 PVK-----VSVTEV--KRHKLETGIRLDSEYGL--CGK--IAYDYNNFNKGYIGSVW 335
 DB 393 PVSKSGSHTVAYMNVGAPAGMNAAVRSTVRIGLIQGNRLVLVVDHGFEGELAKQIEAG 452
 QY 336 DMDKYETTLAAGISQPRNRYGNWTSNVSYNRSTTONLEKRAFS-----GGIYW 384
 DB 453 -----SVVGG-WTGGGSKLGTKRTLPKKSFQISANITKENIQLVI 494

```
QY 385 VRDAGIDARLGAELAEGRK-----IPGSDIDLGNHSA--TMLT- 422
Db 495 I---GGFEATYTGLEMEGRKQFDELCPFVVVVPATVSNVNVGSDFSVGDATLNTICT 551
QY 423 -----ASMKRLNNVLRHPENGYLDGKIGTGLG---TFLSSTALIRTSARAGYFFT 471
Db 552 CDRIKQSAAGTKRRVF-----IETMGCGCYLATMAGLAAGADAAYIFE 596
QY 472 PENKKGTCFIIR---GQAGYTVARDNADVPGLMFRSGGASSVRGYELDSI-GLAPNG- 526
Db 597 E-----PFTIRDLQANVHLVOKWKTTVKRGVLNRNEKCN--ENYTTDFIFNLSEEG 648
QY 527 SVLPERALLVGSLEY---QLPPTRTLGSAGVFDMDGDAANFKRMKLKH----- 571
Db 649 GIFDSRKNVLGHMOOGSGSTPPDRNFA-----TKMGAKAMNMWSGKIKESYRNGRIFANTP 704
QY 572 --GSGLGVR-----WFSPLAPSFDFAYGHSKDKIRWHISL 605
Db 705 DSGCVLGMRRRALVFPVLAELKDQDTDFHRIPKQWMLKL 744

RESULT 9
K6PF CANFA
ID K6PF_CANFA STANDARD; PRT; 781 AA.
AC P52784;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM OR M-PFK.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96194913; PubMed=8654960;
RA Smith B.F., Henthorn P.S., Rajpurohit Y., Stedman H., Wolfe J.H.,
RA Patterson D.F., Giger U.;
RA "A cDNA encoding canine muscle-type phosphofructokinase.";
RL Gene 168:275-276(1996).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -1- COFACTOR: MAGNESIUM.
CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
CC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U25183; AAC48615.1; -.
DR HSP52; P00512; 3PKF.
DR InterPro; IPR000023; Ppfrckinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PHERCTKINASE..
DR ProDom; PD000707; Ppfrckinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
KW Kinase; transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INIT_MET 0 0
FT REPEAT 1 403 APPROXIMATE.
FT
```

```
FT REPEAT 404 781 APPROXIMATE.
FT MOD_RES 776 776 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 781 AA; 85429 MW; 15CA0B45438462EE CRC64;

Query Match 4.1%; Score 131; DB 1; Length 781;
Best Local Similarity 19.9%; Pred. No. 0.55;
Matches 129; Conservative 73; Mismatches 236; Indels 210; Gaps 31;

QY 81 EVIDKQEQTGFELAEAPDNVKTMLRSKYFSSKVSILTEKDGAYTVHTTTPGPRTKIANVGVA 140
Db 186 EIVDALTTTAQSHORFTVLEVMGRHCGYALVLSLSCGADWVFIECPDP----- 235
QY 141 ILGDILSDGNLAEYRNALENMQQVPGSDFDQDSWNSKTSVL-----GAYTRKGYPLA-- 194
Db 236 -----DDWEELHRLSELSETRTGRSLNLIIVAEISADKNGKPTSE 276
QY 195 -----KLGNTRAAVNDTATVDLNVVVDUSGRPIAGDFEITGTORYEQEIVSGLARF 246
Db 277 EIKELVVKRLGY-----DTRVTYVLGHVQGGTSPAFD--RILGS-RMGVEAVMALLEG 326
QY 247 QPGTPYDLDLL-----LDFQALQNGHYSGASVOADF--RLQG----- 284
Db 327 TPDTPACVVSLSGNAQVRLPLMECVQVTKDVTKAMNDRKFDEAMKLGSRFNMNVEYKL 386
QY 285 -DRVVKVSVTEVKRHK-----LETGRLDSEYGL--GGR--TAYDYNLNLPNK 327
Db 387 LAHIRPPVSKTSATMHTVAVMNVGAPAGMNAAVRSTVRIGLTQGNRVLVVHDGPEGLAK 446
QY 328 GYIGSVVMDKYEYTTLAAGISQPRNYRGNTWTNVSYNRSSTTONLEKRAFS----- 379
Db 447 GQIEEAGW-----SYVGG-WTGGGSKLGTKRTLPKSKFEQISANITK 488
QY 380 ---GGTWYVDRAGIDARLGAELAEGRK-----IPGSDIDLGNH 417
Db 489 FNIQGLIII---GGFEATYTGLEMEGRKQFDELCPFVVVVPATVSNVNVGSDFSVGDADT 545
QY 418 A--TMLT-----ASMKRLNNVLRHPENGYLDGKIGTGLG---TFLSSTALIRTS 463
Db 546 ALANTICTCDRIKQSAAGTKRRVF-----IETMGCGCYLATMAGLAAG 590
QY 464 ARAGYFTEFPENKKGTCFIIR---GQAGYTVARDNADVPGLMFRSGGASSVRGYELDSI- 519
Db 591 ADAAYIFE-----PFTIRDLQANVHLVOKWKTTVKRGVLNRNEKCN--ENYTTDFIF 642
QY 520 GLAPNG-SVLPERALLVGSLEY---QLPPTRTLGSAGVFDMDGDAANFKRMKLKH----- 571
Db 643 NLYSEEGKGFDSRKNVLGHMOOGSGSTPPDRNFA-----TKMGAKAMNMWSGKIKESYRN 698
QY 572 -----GSGLGVR-----WFSPLAPSFDFAYGHSKDKIRWHISL 605
Db 699 GRIFANTPDSGCVLGMRRRALVFPVLAELKDQDTDFHRIPKQWMLKL 746

RESULT 10
K6PF_RABIT
ID K6PF_RABIT STANDARD; PRT; 779 AA.
AC P00511;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN
SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=87166033; PubMed=2951385;
RA Lee C.P., Kao M.C., French B.A., Putney S.D., Chang S.H.;
RA "The rabbit muscle phosphofructokinase gene. Implications for protein
RT
```

RT structure, function, and tissue specificity.";
RN J. Biol. Chem. 262:4195-4199(1987).
RP SEQUENCE.
RX MEDLINE=84219739; PubMed=6233492;
RT Poorman R.A., Randolph A., Kemp R.G., Heinrikson R.L.;
RA "Evolution of phosphofructokinase -- gene duplication and creation of
RT new effector sites.";
RL Nature 309:467-469(1984).
CC 1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate -> ADP + D-
CC fructose 1,6-bisphosphate.
CC 1- COFACTOR: MAGNESIUM.
CC 1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC 1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC 1- SUBUNIT: HOMOTETRAMER.
CC 1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M14477; AAA31441.1; .
DR EMBL; M14456; AAA31441.1; JOINED.
DR EMBL; M14457; AAA31441.1; JOINED.
DR EMBL; M14458; AAA31441.1; JOINED.
DR EMBL; M14459; AAA31441.1; JOINED.
DR EMBL; M14460; AAA31441.1; JOINED.
DR EMBL; M14461; AAA31441.1; JOINED.
DR EMBL; M14462; AAA31441.1; JOINED.
DR EMBL; M14463; AAA31441.1; JOINED.
DR EMBL; M14464; AAA31441.1; JOINED.
DR EMBL; M14465; AAA31441.1; JOINED.
DR EMBL; M14466; AAA31441.1; JOINED.
DR EMBL; M14467; AAA31441.1; JOINED.
DR EMBL; M14468; AAA31441.1; JOINED.
DR EMBL; M14469; AAA31441.1; JOINED.
DR EMBL; M14470; AAA31441.1; JOINED.
DR EMBL; M14471; AAA31441.1; JOINED.
DR EMBL; M14472; AAA31441.1; JOINED.
DR EMBL; M14473; AAA31441.1; JOINED.
DR EMBL; M14474; AAA31441.1; JOINED.
DR EMBL; M14475; AAA31441.1; JOINED.
DR EMBL; M14476; AAA31441.1; JOINED.
DR PIR; A00604; KIRBF.
DR PIR; A26550; A26550.
DR HSSP; P00512; 3PKX.
DR InterPro; IPR000023; PpfFructokinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PFRFRCTKINASE.
DR PROSITE; PD000707; PpfFructokinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
KW Phosphorylation; Magnesium; Multigene family.
FT INIT_MET 0 . . . 0
FT REPEAT 1 401 APPROXIMATE.
FT REPEAT 402 779 APPROXIMATE.
FT MOD_RES 774 774 PHOSPHORYLATION.
FT CONFLICT 268 268 R -> S (IN REF. 2).
FT CONFLICT 479 507 MISSING (IN REF. 2).
FT CONFLICT 558 558 S -> I (IN REF. 2).
FT CONFLICT 565 565 MISSING (IN REF. 2).
SQ SEQUENCE 779 AA; 85072 MW; 9608E7C7B5D3C8A9 CRC64;
Query Match 4.0%; Score 128; DB 1; Length 779;
Best Local Similarity 20.0%; Pred. No. 0.86;
Matches 128; Conservative 72; Mismatches 245; Indels 194; Gaps 29;
QY 81 EYLDKEQTGFLAEAPDNVATMLRSKGYFSSKVSUTEKDCAYTVHITGPRTKIANVQVA 140
DB 186 EIVDAITTTAASHQRTVELEVMGRHCGYLALVTSLS--CGADWVFIPSCPPD----- 235
QY 141 ILGDIISDGNLAEEYRNALENWQPVGSDFDODSNENSKTSVL-----GAVTRKGYPLAKL 196
DB 236 -----DNWEDHLCLRLSETRTGRSLNIIIVAEGAIDRNGKPTISE 276
QY 197 GNTRAAVNP--DTATVDLVNVDGSRPTARCFETITGTORYPEQIVSGLARFPQCPYD 253
DB 277 GVKDLVRLRGYDTRVTVLGHVORGTPSAFD--RLTGS-RMGVEAVWALLEGTPTDPAC 333
QY 254 LDLL-----LDFQALEQNGHYSCASVQADFD-----RLQGDVRVPKVSVEVKRH---- 299
DB 334 VVSLSGNQAVRLPLMECVQVTKVAMDEKREDEAMKLRGSRFNNNEVYKLLAHIRPP 393
QY 300 -----KLETGIRLDSYGL--GGK--IAYDYNNLPNKGYSVVVD 336
DB 394 APKSGSYTVAVNVCAPACAAANAAVSTVRIGLIQGNRVLVVDHGFEGPAKQIIEAGW- 452
QY 337 MDKYETTLAAGISOPRNYRGNVWTSNVSYNRSTQNLKRAFS-----GGIYV 385
DB 453 -----SYVGG-WTGGGSKLGSKRTLPKKSPEQISANITFNIQGLVII 495
QY 386 RDRAGIDARLGAELAEGRK-----IPGSDIDLGNSHA--TMLT-- 422
DB 496 ---GFEAYTGGLELMGRKQFDELCPFVVPATVSNVNVPGSDFSGADTALNTICTTC 552
QY 423 -----ASMKRQLLNVLHPENGHYLDGKIGITLG---TFLSSTALRTSARAGYFFTP 472
DB 553 DRIKQSAAGTKRRVF-----IETMGYGYGLATWAGLAAGADAAYIPEE 597
QY 473 ENKKGITFIIR--GOAGYTVARDNADVPGLMFRSGGASSVGRGYELDSI-GLAGPNG-S 527
DB 598 -----PTIRDLOANVEHLVQRMKTTVRKGLVRNEKCN--ENTYTFIFMLYSEEGK 649
QY 528 VLPERALLVGSLEY---OLPPTRTLSCAVFHDMDGDAANFKRMKLKH----- 571
DB 650 IPDSRKNVLGHMQGGSPTPFDRNFA----TKGAKAMNMAGKIKESYRNGRIFANTPD 705
QY 572 -GSGIGVNR---WFSPLAPFSFDIAYGHSKKIRWHISL 605
DB 706 SCVVLGMRKRALVPQVTELQNTDFEHRIPKEQWNLKL 744
RESULT 11
YEBT_HAETIN
ID YEET_HAETIN STANDARD; PRT; 881 AA.
AC P44288;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein H11672.
GN H11672.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klerlagave A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Frithman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";


```

RESULT 13
K6PF_RAT
ID K6PF_RAT STANDARD; PRT; 779 AA.
AC P47858; Q63736;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6-phosphofructokinase, muscle type (EC 2.7.1.11) (Phosphofructokinase
DE 1) (Phosphohexokinase) (Phosphofructo-1-kinase isozyme A) (PFK-A).
GN PFKM OR PFK-M.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=pancreatic islets;
RX MEDLINE=96550468; PubMed=8764833;
RT Ma Z., Ramanadham S., Turk J., Kempe K., Hu Z., Ladenson J.;
RT "Characterization of expression of phosphofructokinase isoforms in
RT isolated rat pancreatic islets and purified beta cells and cloning
RT and expression of the rat phosphofructokinase-A isoform.";
RL Biochim. Biophys. Acta 1308:151-163(1996).
RN [2]
RP REVISIONS TO 162-169 AND 493-497.
RA Ma Z.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 7-212 FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Nakajima H.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -!- COFACTOR: MAGNESIUM.
CC -!- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
CC -!- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U25651; AAC52786.1; -
CC DR EMBL; D21869; BAA21013.1; -
CC DR HSSP; P00512; 3PFK.
CC DR InterPro; IPR000023; Ppfuckkinase.
CC DR Pfam; PF00365; PFK; 2.
CC DR PRINTS; PR00476; PHFRCTKINASE.
CC DR ProDom; PD000707; Ppfuckkinase; 2.
CC DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
CC KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;
CC Phosphorylation; Magnesium; Multigene family.
CC INIT_MET 0
CC BY SIMILARITY.
CC FT MOD_RES 774 774 PHOSPHORYLATION (BY SIMILARITY).
CC FT CONFLICT 32 32 T -> A (IN REF. 3).
CC FT CONFLICT 45 45 L -> A (IN REF. 3).
CC FT CONFLICT 149 149 R -> A (IN REF. 3).
CC FT CONFLICT 159 159 F -> G (IN REF. 3).
CC FT CONFLICT 179 179 S -> F (IN REF. 3).
CC FT SEQUENCE 779 AA; 85428 MW; F4C92DDE47D53788 CRC64;
CC
CC Query Match 3.8%; Score 120; DB 1; Length 779;
CC Best Local Similarity 20.3%; Pred. No. 2.8;
CC Matches 129; Conservative 77; Mismatches 247; Indels 188; Gaps 31;

```


RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.*;

RL Nature 406:959-964(2000).
CC -1- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON

CC STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.

CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; L10210; AAA5819.1; -
DR EMBL; U07359; AAB6019.1; -
DR EMBL; AE004666; AAG05786.1; -
DR PIR; A40601; A40601.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
FT SITE 798 815 TONB C-TERMINAL BOX.
FT CONFLICT 716 715 Y -> F (IN REF. 1).
FT CONFLICT 745 746 MISSING (IN REF. 1).
FT CONFLICT 750 750 S -> R (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; AIIAE4A9A290F35 CRC64;

Query Match 3.7%; Score 118.5; DB 1; Length 815;
Best Local Similarity 19.5%; Pred. No. 3.8;
Matches 112; Conservative 68; Mismatches 192; Indels 201; Gaps 30;

QY 66 DMVEHPLTQQEEVLDEKTEGFLAEAP-----DNVKTMLRSKGYFSSKV--SITE 117
DB 186 DVVHRHTPGIT--VSAYIDRNNYARGFSINFDYDGPSTARNVGSAGNTLSDMAI 242
QY 118 KDAYTVHTPGPRTKIANGVAILGDILSDGNLAIEYRNALFNQQPVGS-----DFDQ 172
DB 243 YDRVEVLKATGLLTGAGSLGATV-----NLIR-----KKPTHEFGKHWELGA 285
QY 173 DSWEN--SKTSVLGAVTRKGYPLAKLGNTRAA-----PLTESGNVGRVAAVQDKHFMDFHBERKTSVYVYGILEF 338
DB 286 GSWDNRSELDVSG-----SCRTAFSGDFETTGQRYPEQIVISGLARFQPGTPYD 253
QY 203 -VNPDTATVDLNVVVD-----LTVGADYQDNDPKPGSGSGFPLFDSQGNRNDVSRFNNGAKWSWEQYT 395
DB 339 DLNPDPTM---LTVGADYQDNDPKPGSGSGFPLFDSQGNRNDVSRFNNGAKWSWEQYT 395
QY 254 LDLLLDFOALEQ-----NGHYSGASVQADFRLQGD-----RVPVKYSV 293
DB 396 RTVFANLEHNFAWGKGVQVLDHKLINGIH-----APLGATMGDWPAPDNSAKIVQAQYKT 449
QY 294 TEVRKHLE---TG-----IRLDSYGLGKAIAYDYNLFNKGYISGVYDMDKYETTLAA 346
DB 450 GETKNSLDIYLTGFPFLGREHLYVGTSSPSHWE-----GKSYWNLNRYDNTDITDD 502
QY 347 GTS-----OPRNY-----RGNYWTSTNSVSNSTTONLEKRAFSGGIWYVDR 388
DB 503 FFWDDGDKGPDWGTIPSOYIDDKTRQUGSWTAR--FNVTDLLN-----FLGG----- 549
QY 389 AGIDAR---LGAELFAERKIP--GSDIDLGNSHAT-----MLTASWKRQLLNLYLH 435
DB 550 RVVDYRVRTGLNPTIRSGRFTPYVGVAYVLDNTYSVYASVTDIFMPQDSWYRDNSSNKLLE 609
QY 436 PE-----NGHYLDGKIGTTLGTFLSLTALIRTSARAGYFFTPENKLGTFPIIRG-- 484

Db 610 PDEQNYEIGIKGEYLDGRINTSLAYF-EIHEENRAEDALYNSKPTNPAL-TTAYKGIK 667
QY 485 -----QAGYT--VARDNA 495
DB 668 AKTKGYEAEISGELAPGWQVQAGYTHKIIRDDSD 700
RESULT 16
120K_RICKI STANDARD; PRT: 1300 AA.
ID 120K_RICKI AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R.
RX MEDLINE=90136087; PubMed=25154118;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT *Cloning, expression and sequence analysis of the gene encoding the
RT 120 kDa surface-exposed protein of Rickettsia rickettsii.*;
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; X16353; CAA34402.1; -
DR PIR; S07575; S07575.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).

```
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match 3.7%; Score 116.5; DB 1; Length 1300;
Best Local Similarity 19.6%; Pred. No. 9.7;
Matches 117; Conservative 92; Mismatches 248; Indels 139; Gaps 24;

QY 72 LPIITQQQEEVLD-----KEQTGLAEAPDNVKTMLRSKYFSSKVSLETKDGAYTVH 125
DB 219 IKLTSTQNNIVVDFDLATDQTGVDASSLTNAQTL-----TINGKIG----- 262
QY 126 ITPGPRTK-IANVGVAIIGDILSDGNLAEYRYNALENWQPVGSD-----PDQDSWENSKT 180
DB 263 -TIGANNKTLGQFNIGSSKTVLSNGVA---INEL-----VIGDCGAVQFAHDTYLITRT 313
QY 181 S-----VLCAVTRKGYPLAKLGNTRAAVNP-----DTATVDLVNVVDSGRPIAFGD 226
DB 314 TNAAGQCKIIFNPVYVNGTTLAAGTNLGSATNPLAEINFGSKGVNVDVTLNVGEGVNLVA 373
QY 227 FEITGTQRY-----PEQIVSGLARFQPTPYDLDDL-----L 258
DB 374 TNLITTDANVGSFEVFNAGGTNI VSGTVGGQGNKFNVALENGTTVKFLGNATFNGNTTI 433
QY 259 DFQQAALQNGHYSGASV-QADFRLQGRDVPVKVSVTEVYKRRKLETGIRLDSEYGLGCKI 317
DB 434 AANSTLQIGGNTYADCVASADGTGIVEFVNTGPTVTLNQAAPVNAKQITVSGPGNVV 493
QY 318 AYDYNN-----LFNKGYIGSVWMDMKYETTLAAGISQPRNRYGNWTSNVS 365
DB 494 INEIGNAGNHHCVAVTDITAFENSSILGAVP-----LPRGI--PFNDAGNTMPLTI-- 541
QY 366 NSTTQNLKRAFSGGIYVDRAGIDARLCAEFLAEGRKIPGSD-----IDLGNSHATML 421
DB 542 -KSTVGNKTAKGFDVPSVVV---LGVDSD-----VIADGVQVQDNNIVGLGSDNGIIV 592
QY 422 TASMKRQLLNVLHPENGHYLDGKIGTTLGTLFSLSTALIRTSARAGYFETPENKLGTFI 481
DB 593 NATTLVAGISTLNNQGTVTLSGGVPNTPGTVYGLGTGIGASKFKQVTTTDDYNLGNII 652
QY 482 IRQAGYTVARDNADVPGLMFRSGGASSVGYELDSIGLAGPNS-----VLPERA 533
DB 653 -----ATNATINDGVTVTTGGIAGI-GFD-GKITLGSVNGNGNRYFADGILSNST 700
QY 534 LLVGSLEYQLPPTRLTSGAVFHDMDGA---AANFKRMKLKHSGSLGRWFSPLEAF 586
DB 701 SMITGTTKANNGTVTVLGNFAFVNGIGSDTVPASVRFTGSDSGAGLQGNISQVIDF 756

RESULT 17
OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOMPb)
DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=92167802; PubMed=1724278;
RA Gilmore R.D. Jr., Cleptak W. Jr., Pollicastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (rOMP B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor.";
RL Mol. Microbiol. 5:2361-2370(1991).
RN [2]
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16353; CAA34403.1; -
DR InterPro; IPR003858; rOMPb_rOMPb.
DR Pfam; PF02708; rOMPb_rOMPb; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1334 1654 32 KDa BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

Query Match 3.7%; Score 116.5; DB 1; Length 1654;
Best Local Similarity 19.6%; Pred. No. 14;
Matches 117; Conservative 92; Mismatches 248; Indels 139; Gaps 24;

QY 72 LPLITQQQEEVLD-----KEQTGLAEAPDNVKTMLRSKYFSSKVSLETKDGAYTVH 125
DB 573 IKLTSTQNNIVVDFDLATDQTGVDASSLTNAQTL-----TINGKIG----- 616
QY 126 ITPGPRTK-IANVGVAIIGDILSDGNLAEYRYNALENWQPVGSD-----FDQDSWENSKT 180
DB 617 -TIGANNKTLGQFNIGSSKTVLSNGVA---INEL-----VIGDCGAVQFAHDTYLITRT 667
QY 181 S-----VLCAVTRKGYPLAKLGNTRAAVNP-----DTATVDLVNVVDSGRPIAFGD 226
DB 668 TNAAGQCKIIFNPVYVNGTTLAAGTNLGSATNPLAEINFGSKGVNVDVTLNVGEGVNLVA 727
QY 227 FEITGTQRY-----PEQIVSGLARFQPTPYDLDDL-----L 258
DB 728 TNLITTDANVGSFEVFNAGGTNI VSGTVGGQGNKFNVALENGTTVKFLGNATFNGNTTI 787
QY 259 DFQQAALQNGHYSGASV-QADFRLQGRDVPVKVSVTEVYKRRKLETGIRLDSEYGLGCKI 317
DB 788 AANSTLQIGGNTYADCVASADGTGIVEFVNTGPTVTLNQAAPVNAKQITVSGPGNVV 847
QY 318 AYDYNN-----LFNKGYIGSVWMDMKYETTLAAGISQPRNRYGNWTSNVS 365
DB 848 INEIGNAGNHHCVAVTDITAFENSSILGAVP-----LPRGI--PFNDAGNTMPLTI-- 895
QY 366 NSTTQNLKRAFSGGIYVDRAGIDARLCAEFLAEGRKIPGSD-----IDLGNSHATML 421
DB 896 -KSTVGNKTAKGFDVPSVVV---LGVDSD-----VIADGVQVQDNNIVGLGSDNGIIV 946
QY 422 TASMKRQLLNVLHPENGHYLDGKIGTTLGTLFSLSTALIRTSARAGYFETPENKLGTFI 481
```


DOMAIN 629 820
B DOMAIN
N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 70
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 539
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 961
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 452
FT TURN 453
FT TURN 466
FT TURN 478
FT STRAND 480
FT STRAND 481
FT STRAND 486
FT STRAND 492
FT STRAND 493
FT STRAND 497
FT STRAND 503
FT STRAND 504
FT STRAND 505
FT STRAND 509
FT STRAND 514
FT STRAND 515
FT STRAND 518
FT STRAND 523
FT STRAND 528
FT STRAND 533
FT STRAND 539
FT STRAND 541
FT STRAND 542
FT STRAND 552
FT STRAND 553
FT STRAND 557
FT STRAND 561
FT STRAND 563
FT STRAND 565
FT STRAND 569
FT STRAND 573
FT STRAND 577
FT STRAND 578
FT STRAND 582
FT STRAND 584
FT STRAND 588
FT STRAND 589
FT STRAND 591
FT STRAND 593
FT STRAND 600
FT STRAND 610
FT STRAND 618
FT STRAND 628
FT STRAND 629
FT STRAND 635
FT STRAND 636
FT STRAND 637
FT STRAND 648
FT STRAND 651
FT STRAND 657
FT STRAND 660
FT STRAND 667
FT STRAND 669
FT STRAND 670
FT STRAND 671
FT STRAND 672
FT STRAND 674
FT STRAND 677
FT STRAND 680
FT STRAND 681
FT STRAND 682
FT STRAND 684
FT STRAND 687
FT STRAND 698
FT STRAND 702
FT STRAND 703
FT STRAND 705
FT STRAND 707
FT STRAND 708
FT STRAND 709
FT STRAND 710
FT STRAND 711
FT STRAND 712
FT STRAND 713
FT STRAND 714
FT STRAND 715
FT STRAND 716
FT STRAND 717
FT STRAND 718
FT STRAND 719
FT STRAND 720
FT STRAND 721
FT STRAND 722
FT STRAND 723
FT STRAND 724
FT STRAND 725
FT STRAND 726
FT STRAND 727
FT STRAND 728
FT STRAND 729
FT STRAND 730
FT STRAND 731
FT STRAND 732
FT STRAND 733
FT STRAND 734
FT STRAND 735
FT STRAND 736
FT STRAND 737
FT STRAND 738
FT STRAND 739
FT STRAND 740
FT STRAND 741
FT STRAND 742
FT STRAND 743
FT STRAND 744
FT STRAND 745
FT STRAND 746
FT STRAND 747
FT STRAND 748
FT STRAND 749
FT STRAND 750
FT STRAND 751
FT STRAND 752
FT STRAND 753
FT STRAND 754
FT STRAND 755
FT STRAND 756
FT STRAND 757
FT STRAND 758
FT STRAND 759
FT STRAND 760
FT STRAND 761
FT STRAND 762
FT STRAND 763
FT STRAND 764
FT STRAND 765
FT STRAND 766
FT STRAND 767
FT STRAND 768
FT STRAND 769
FT STRAND 770
FT STRAND 771
FT STRAND 772
FT STRAND 773
FT STRAND 774
FT STRAND 775
FT STRAND 776
FT STRAND 777
FT STRAND 778
FT STRAND 779
FT STRAND 780
FT STRAND 781
FT STRAND 782
FT STRAND 783
FT STRAND 784
FT STRAND 785
FT STRAND 786
FT STRAND 787
FT STRAND 788
FT STRAND 789
FT STRAND 790
FT STRAND 791
FT STRAND 792
FT STRAND 793
FT STRAND 794
FT STRAND 795
FT STRAND 796
FT STRAND 797
FT STRAND 798
FT STRAND 799
FT STRAND 800
FT STRAND 801
FT STRAND 802
FT STRAND 803
FT STRAND 804
FT STRAND 805
FT STRAND 806
FT STRAND 807
FT STRAND 808
FT STRAND 809
FT STRAND 810
FT STRAND 811
FT STRAND 812
FT STRAND 813
FT STRAND 814
FT STRAND 815
FT STRAND 816
FT STRAND 817
FT STRAND 818
FT STRAND 819
FT STRAND 820
FT STRAND 821
FT STRAND 822
FT STRAND 823
FT STRAND 824
FT STRAND 825
FT STRAND 826
FT STRAND 827
FT STRAND 828
FT STRAND 829
FT STRAND 830
FT STRAND 831
FT STRAND 832
FT STRAND 833
FT STRAND 834
FT STRAND 835
FT STRAND 836
FT STRAND 837
FT STRAND 838
FT STRAND 839
FT STRAND 840
FT STRAND 841
FT STRAND 842
FT STRAND 843
FT STRAND 844
FT STRAND 845
FT STRAND 846
FT STRAND 847
FT STRAND 848
FT STRAND 849
FT STRAND 850
FT STRAND 851
FT STRAND 852
FT STRAND 853
FT STRAND 854
FT STRAND 855
FT STRAND 856
FT STRAND 857
FT STRAND 858
FT STRAND 859
FT STRAND 860
FT STRAND 861
FT STRAND 862
FT STRAND 863
FT STRAND 864
FT STRAND 865
FT STRAND 866
FT STRAND 867
FT STRAND 868
FT STRAND 869
FT STRAND 870
FT STRAND 871
FT STRAND 872
FT STRAND 873
FT STRAND 874
FT STRAND 875
FT STRAND 876
FT STRAND 877
FT STRAND 878
FT STRAND 879
FT STRAND 880
FT STRAND 881
FT STRAND 882
FT STRAND 883
FT STRAND 884
FT STRAND 885
FT STRAND 886
FT STRAND 887
FT STRAND 888
FT STRAND 889
FT STRAND 890
FT STRAND 891
FT STRAND 892
FT STRAND 893
FT STRAND 894
FT STRAND 895
FT STRAND 896
FT STRAND 897
FT STRAND 898
FT STRAND 899
FT STRAND 900
FT STRAND 901
FT STRAND 902
FT STRAND 903
FT STRAND 904
FT STRAND 905
FT STRAND 906
FT STRAND 907
FT STRAND 908
FT STRAND 909
FT STRAND 910
FT STRAND 911
FT STRAND 912
FT STRAND 913
FT STRAND 914
FT STRAND 915
FT STRAND 916
FT STRAND 917
FT STRAND 918
FT STRAND 919
FT STRAND 920
FT STRAND 921
FT STRAND 922
FT STRAND 923
FT STRAND 924
FT STRAND 925
FT STRAND 926
FT STRAND 927
FT STRAND 928
FT STRAND 929
FT STRAND 930
FT STRAND 931
FT STRAND 932
FT STRAND 933
FT STRAND 934
FT STRAND 935
FT STRAND 936
FT STRAND 937
FT STRAND 938
FT STRAND 939
FT STRAND 940
FT STRAND 941
FT STRAND 942
FT STRAND 943
FT STRAND 944
FT STRAND 945
FT STRAND 946
FT STRAND 947
FT STRAND 948
FT STRAND 949
FT STRAND 950
FT STRAND 951
FT STRAND 952
FT STRAND 953
FT STRAND 954
FT STRAND 955
FT TURN 956

FT TURN 958 961
FT STRAND 966 972
FT TURN 974 976
FT STRAND 979 986
FT TURN 988 989
FT STRAND 991 992
FT STRAND 995 997
FT STRAND 1001 1002
SQ SEQUENCE 1018 AA; 113351 MW; B11B3500A7F0F872 CRC64;

Query Match 3.6%; Score 113.5; DB 1; Length 1018;
Best Local Similarity 20.9%; Pred. No. 11;
Matches 134; Conservative 78; Mismatches 235; Indels 195; Gaps 33;

Qy 69 EEHLPLTTOOQEEVLDKEQTGLAEAPDNVKTMLRSKGYESSKVSLETKDGAVTVHTP 128
Db 92 DKNLALTEEEKMEVI--RNRGIPADVLAKEHFEKHAHESLKDQIYFAVKLSTKV-- 147
Qy 129 GPRTKIANGVAILGD--ILSDGNLAEEYRNALEN-----WOQPV-----GS 168
Db 148 NKFAKIMNLRQSVYVDLKLTDGKLYEGKHIPVSNISAGENHVQIPLMAQEEILSSAS 207
Qy 169 DFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDPE 228
Db 208 DF-----KTAM---VSKSKPKQATAMHGA-----IBIIDS---FASPCDN 243
Qy 229 ITGT-----ORYPEQIVSGL--ARFQPGTPYDLDLDFQOALEQ-----NGHYSAS 274
Db 244 IVGAMLLVDVYHTNPENAVRSIFVAPRGGRIPI---RVVTFPTTIVQIEPDMSRFOLLS 300
Qy 275 VOADFDRLQG--DRVPKVSV-----TEVKKRHKLETGIRLDS---EYGLGKIAYD 320
Db 301 TTTNGDFVQOKOLAMVKNVACAVGLTSSYPTPLLESGLOKDRGLIVEY--FORMSY- 357
Qy 321 YYNLFNKGYIGSVVMDMDKYETTTLAAGISOPRN---YRGNWTSNVSYNRSTTONLEKRA 377
Db 358 -----VAHNTNQPOEKDLLEGNFSDIKRSRLEKVSSTKAQ 394
Qy 378 FSGGIWYVRDAGIDARLGAELAEGRKIPGS-----DIDLGNSHATMLTASW 425
Db 395 FVSGRTPKYDIIGAGSSSEELSEE--KIOGKAKQVDARLORIDPOYNEVQAOETNLF 452
Qy 426 KRQLLNVLHPENGHYLDGKI-----GPTL-----GTFLLSTALIR 461
Db 453 KLS-LDDVETPK-GSMLDLKISQSKIALPKNTYGGTILRSDLLANFLTEGNFRASVDLQR 510
Qy 462 TSARAGYF-----FTPENKKG-----TFIRGOAG---YTVARDNADV-----PSG 500
Db 511 THRIKGMKMWATVGPENTGIALACAMNSSINGRASSDIYICSDQCELANPACTRAMT 570
Qy 501 LMFRRSGGASSV-----RGYELDSIGLAGNGSVLPERALLVG-----S 538
Db 571 MSFNPNCSDAWSLEFLKRTGFHCDICTGTATPMQDVQVTIDWFISSQECVPRTYCV 630
Qy 539 LEYOLPF--TRTLSGAVFDMGDAAAANFKRMKLKHGSGLGV 578
Db 631 LNPQNPFVLNRMWKGKLTFPQ--GTSRSVKRMPLSIGGGAGAK 670

RESULT 21
ID HLYA_PROMI STANDARD; PRT; 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]

SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN-Isolate 477-12;
RX MEDLINE-90170827; PubMed-2407716;
RA Uphoff T.S., Welch R.A.;
RT "Nucleotide sequencing of the proteus mirabilis calcium-independent
RT Serratia marcescens hemolysin genes (shla and shlb).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC DEFINED.
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC REQUIRES HMB FUNCTION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -1- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30186; AA225657.1; -
CC PIR; A35140; A35140.
CC Hemolysis; Toxin; Outer membrane; Signal.
CC SIGNAL 1 29
CC CHAIN 30 1577 HEMOLYSIN.
CC SEQUENCE 1577 AA; 15869 MW; 175975E0C924B2D9 CRC64;
Query Match 3.6%; Score 113.5; DB 1; Length 1577;
Best Local Similarity 21.6%; Pred. No. 20;
Matches 133; Conservative 75; Mismatches 264; Indels 145; Gaps 28;
QY 31 AAGFALF-----KNKPSVESKLPKPFVPRIDTQDSSEIKDMVEHLPLITQOOEVLDRK 86
DB 697 SAGPHITHTNKANTETEQAN-STISANDVLOAN--KDVTFAGSDLTKTAGNASITGD 753
QY 87 OTGFLAE---APDNVKNLRSKGVFS-----SKVSLTEKDGAYTHIITPGPRTKIANVG 138
DB 754 NVAFVSTENKKQTDNTDTTI--SGFSYTGQVDKVG--SKADFQYDKQHTQTEVTRNGSG 810
QY 139 VAILGDILSDGN---LAEYRNLNENWQPPVGSDFDQDSWENSKTSVLGAVTRKGYPLAK 195
DB 811 TEVAGDLTITANKLLHESGASHHVEGRYQESGENIQHVLAVNDSETS-----856
QY 196 LGNTRAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLD 255
DB 857 -----KTDLSLVGIDVGVNLDYSGV---TKPVKKAIEDGVNTTKPGNNTDLT 900
QY 256 LLLDFQOALEQNGHYSGASVADFRLQGDVPVKYSVTEVKRHKLETGIRLDSYGLGG 315
DB 901 KKVTAARDAIAN-----LANLSNETPNVGVGEIKGGGSOQSQDSQAVSTISNAG 951
QY 316 KTAIDYYN-LFNKGVIGSVWMDKYYTTTLAAGISQPRNYRGNWTSNYSNRSTONLE 374
DB 952 KIDIDSNKLLHDQG-----THYQST-QEGIS-----LTANTHTSEATLDKHTTPE 997
QY 375 KRAFGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDTDLGNSHATMLTASWKROLLNN-- 432
DB 998 TK--GGG-----QIGVSTKIGSD-ITVAIKGECQTD-----NALMETKAKSQTSNGD 1044
QY 433 --VLHPENGHY-----LQKGTGTTL-----GTFLLSTALIRTSARAGYFFTPPE 473
DB 1045 ISINVENAHYEGAQFDAQKGTVINAGGDLTLAQATDTHSESQSNVNGSANKLVGTTPPE 1104
QY 474 NKKLGTFTIRGQAGYTVARONADV-----PSGLMFRSGGASSVRGVYELDS---IGLAGPN 525
DB 1105 SKDYGGGFGNAGTTHHSKEQTTAKVGTITGSGQGLNELNAGNINLTQTHLSSQDIALNATN 1164

QY 526 GSVLPERALLVGSLEQLPFTRTLSCGAVPHDMG-----DAAANFKRMKLNKHSGLGVNRF 581
DB 1165 -----KVDLQSSASHTKGNLGGVQAGFGKKMTDDASSV-----NGLG----- 1205
QY 582 PLAPFSPDIAYGSHDKK 598
DB 1206 -----SAQFAIGKQDEK 1217
RESULT 22
YRAJ_ECOLI STANDARD; PRT; 838 AA.
AC P42915;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein yraj precursor.
GN YRAJ OR B3144
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: COULD BE INVOLVED IN THE EXPORT AND ASSEMBLY OF THE
CC PUTATIVE YRAJ FIMBRIAL SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18997; AAA57947.1; ALT_INIT.
CC EMBL; AE000395; AAC76178.1; -
CC EcoGene; EG12775; yraj.
CC InterPro; IPR000015; Fimb_usher.
CC Pfam; PF00577; usher; 1
CC PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 838 HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YRAJ.
FT DISULFID 815 837 POTENTIAL.
SQ SEQUENCE 838 AA; 93616 MW; FD265CC903A269BA CRC64;
Query Match 3.6%; Score 113; DB 1; Length 838;
Best Local Similarity 18.3%; Pred. No. 8.8;
Matches 106; Conservative 63; Mismatches 166; Indels 244; Gaps 28;
QY 141 ILGLDILSDGNLAAYR-----NALENWQPPVGSDFDQDSWENSKTSVLGAVTRKG 190
DB 254 LLGESATSGSIFSSYFTTGVLASDDNNLPSQR---GFAPTVRCIANSSAIVTRQNG 309
QY 191 YPLAKLGNTRAVNPDTATVDLNVVDSGRPIAFGDFEIT-----GTQRYPEQIVSGLAR 245
DB 310 YVIYO-----SNVSAGAFEINDLYPSSNS---GDLVETIESDGTOR-----R 349

QY 246 F-CPGTPYDLDLLDFQALFONGH---YSGASVQADFDRLOGDRVPVKVSVTEV---KR 298
 Db 350 FIQPIYSSLPH-----MORPHLYSATAGRYRADANSKSEFEAEATAIYGLANT 400
 QY 299 HKLETGIRLDSY-----GLGG----- 315
 Db 401 FTYGLGLGSEYIALGIGIGITGLGALGALSMDINRADTQFONHSHFGHYOWRTQYIKDI 460
 QY 316 -----KIADYNYLNFNKGII-----GSVVMDKYE-----TTLAAGIS-----QP 351
 Db 461 PETNTNIAYSYRYTNDGYFSFNEANTRWYDYSROKSEIOPNISQITFDGVSILVASGSQ 520
 QY 352 RNYRNY-----WTSVSNVNRSTQNLEKRAFSGGI-----WYVRD 387
 Db 521 ODYWGNDKRNISVGVSGOOWGVYSLNYQYGRYTDON-NDRALSLSLISPLERWLPRS 579
 QY 388 RAGI-----DARLCAEFLAERKIPG--SDIDLGNHATMLTASHKROL----- 429
 Db 580 RVSQWTSOKDRPTQHEMRLDGSLDGLRLSYLSLQSLDDNNHNSLNASYRSPGYTFS 639
 QY 430 -----LNN-----VLPENGHYLDGKIGTGLTGLFSLST-ALIRTSARAG----- 467
 Db 640 AGYSYGNDSQYNYGVTVGVVHHP-----GVTLSOYLGNAFALIDANGASGVRIQ 690
 QY 468 -----YFFPENKKLCT-----FTIRGOAGYTVARDNA 495
 Db 691 NYPGIATDPFGYAVPYLYTYQENLSVDTQLPDNDVLEQITQFVVPNRGAMVAARNNA 750
 QY 496 DVPGLMFRSGASSVRGYELDSIGLAGPNSVLPRLAL 534
 Db 751 NI-----GYRV-LVTVSDRNGKLPFGAL 773

RESULT 23
 ID VPL_BPCHP STANDARD; PRT; 595 AA.
 AC P19192;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DE 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Protein VPI (ORF1).
 OS Bacteriophage Chpl.
 OC Viruses; ssDNA viruses; Microviridae; Microvirus.
 OX NCBI_TaxID=12367;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
 RX MEDLINE=90111716; PubMed=2607341;
 RA Storey C.C., Lusher M., Richmond S.J.;
 RT "Analysis of the complete nucleotide sequence of Chpl, a phage which
 infects avian Chlamydia psittaci.";
 RL J. Gen. Virol. 70:3381-3390(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D00624; BAA00515.1;
 DR PIR; J00345; J00345.
 KW Structural protein.
 FT INIT_MET 0
 FT SIMILAR 16 73 TO F PROTEIN OF PHAGE PHI-X174, AA
 FT RESIDUES 2-58.
 SQ SEQUENCE 595 AA; 66850 MW; AE71D8024EA967ED CRC64;
 Query Match 3.58; Score 111.5; DB 1; Length 595;
 Best Local Similarity 23.58; Pred. NO. 6.8;
 Matches 107; Conservative 55; Mismatches 214; Indels 79; Gaps 22;
 QY 14 FPFPHAYAAPADLSENKAAGFALFKNKSPTDESVKLKPFPFRIDTQDSSEIKDMVEEHL 73

Db 117 YFVQVTSDDGYSAGNSIYDFGLPTKVANYRH-QVLPRLAYMLIFENYTRDNQESLP 175
 QY 74 LITOOOEELDKDEQTFLAEE---APDNVTKMLRSK--GYFSSKSVLTKDGAYTVHTHP 128
 Db 176 VMTGDADPKVD-PTTGESQEDDAVPVYVYKLMRRNKRKYDYFTSALPCLOKQSPSVGIGITG 234
 QY 129 GPRTKIANYGVAI---LGDILSDG-NLAERYNALENWQOPVGSDFDQDSWENSKTSVLG 184
 Db 235 GDSGRPLVHGLAIRSYLDDSSDDQFSFGVSVNASQKW-----FTAD---GRLTSGMG 284
 QY 185 AV---TRKGYPL-----AKLGNTRAAVNPDTATVDLNVVVVDGSRPIAFGDFEITGTORY 235
 Db 285 SVPVGTGTFNFDIMVVPYSYFCTTVAQTGSPSS-----STPPFVKGDPEV-----Y 331
 QY 236 PQOIVSGLARFOPGPGYFDLDDLLDFQOALEONGHYSG---ASVQADPDRLOGD---RVPV 289
 Db 332 VD-----LAASSVTINSIRNAITLQWFEXSARYSGRYVESVQGHFVHLGDYRAORPI 386
 QY 290 -----KVSVEYKRRHKLETGIRLDSYGLGKIAI---DYNNLFNKGII---GSVYVD 336
 Db 387 YLGGSKSYSVNPVYQNS-----STDSVSPQNLSAVALSTDTKHLFTKSFVHGFVIGL 441
 QY 337 MD-----KYETTLAAGISQPNRYRGNYWTSNVSYNRSTTONLEKRAFSGGIHYVRDRAGI 391
 Db 442 LSNATADLYTQOGLERQNSRFERY-DYWTPTFAHLGEOPVYNKELYCQSDVTMDPSPGSAYN 500
 QY 392 DARLG-AEFLAERKIPGSDIDLGNHATMLTASW 425
 Db 501 DVPPGVOERYAEYRYKPSKVTGLFRSNATGTLDW 535
 RESULT 24
 ID TACY_STRCB STANDARD; PRT; 574 AA.
 AC Q53957;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Streptolysin O precursor (Thiol-activated cytolysin).
 GN SLO
 OS Streptococcus canis.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95102113; PubMed=7803818;
 RA Okumura K., Hara A., Tanaka T., Nichiguchi I., Minamide W.,
 RA Igarashi H., Yutsudo T.;
 RT "Cloning and sequencing the streptolysin O genes of group C and group
 G streptococci.";
 RL DNA Seq. 4:325-328(1994).
 CC -!- FUNCTION: SULFHYDRYL-ACTIVATED TOXIN. IS ABLE TO LYSE CHOLESTEROL
 CONTAINING MEMBRANES. CAN BE REVERSIBLY INACTIVATED BY OXIDATION.
 CC CHOLESTEROL IS THE RECEPTOR FOR THE BINDING OF THESE TOXINS TO
 CC EUKARYOTIC CELL MEMBRANES.
 CC -!- SIMILARITY: BELONGS TO THE THIOL-ACTIVATED CYTOLYSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D16825; BAA04105.1;
 DR HSP; P19995; LPFO.
 DR InterPro; IPR001869; Thiol_cytolysin.
 DR Pfam; PF01289; Thiol_cytolysin; 1.
 DR PRINTS; PR01400; TACYTOLYSIN
 DR ProDom; PD007062; Thiol_cytolysin; 1.


```
OY 316 KIAYDYNNLEN-KG-YIGSVVMDKXYETTLAAGISOPRNYRCNYWTSNYSNRSTTONL 373
DB 366 RVANGQVOIFSGRGTAGT-----DSSLTNIRAPLOFSGPALTASLOGSGPITNS 417
OY 374 EKRAF-----SGGIWYVRDRAGIDARLCAEF-----LAERKIPGSDIDLG-----NS 416
DB 418 NNGTFLGSLIGPMVMVDQNRQVNPAGLVFQGNLVPNLADPLAIDSKISLSLGPGLTQ 477
OY 417 HATMLTASMKRLLNNVLPENGHYLDGKIGTTLGTFSLSTALIRTSARAGYFFETPENKK 476
DB 478 ASNALTLS-----LGNLEPSN-QAVAIKAGRL-RFESSOALESSLTVNGLT----- 525
OY 477 LGTFIIRGOAGYTV-ARDN---ADVPGLMFRSGGASSVRGYELDSIGLAGPNSVILP-- 530
DB 526 LDTVIRPNLGDGLVEVRDNRIIVKLGANLRFENGAVTA-----CTVNPSAPEAPPT 576
OY 531 ---ERALLVGSLEYQLPFTTLTSLGAVFHD-----MGDA-AANFKRMKLGKSGGLVGR 578
DB 577 LTAEPPLRASNSHLQLSLS---EGLVVHNNALALQDGMENVQHGLTLRVGSGLQMR 631

RESULT 26
YHJG_ECOLI
ID YHJG_ECOLI STANDARD; PRT; 691 AA.
AC P37645; P37644;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Hypothetical protein yhjg.
GN YHJG OR B3524.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RL region from 76.0 to 81.5 minutes."
RN Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RP REVISIONS.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RC Science 277:1453-1474(1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Query Match 3.5%; Score 110.5; DB 1; Length 691;
Best Local Similarity 20.0%; Pred. No. 9.8;
Matches 96; Conservative 53; Mismatches 143; Indels 187; Gaps 23;
```

```
OY 161 NWQPVGSDPDSDNSWENKTSVLGAVTRKGYPLAKLGN-----RAAVNPDTATVNLNV 215
DB 144 NMTFNLANDKDA--NAKPSAMS-----FRLDNLFDOGRITADDKVSKADLEIF 192
OY 216 VDS-GRPIAFGDFEITQRYPEQIVSGLARFQPGTYPDLDLILDFQOALEONGHYS GAS 274
DB 193 VDPLGKPLPFS--EVTGSK-----GKA-----DKEKVG DYVFLGKAGQRYNG-- 232
OY 275 VOADPDRLOGRVPVKVSVTEVRKHKLETGIRLDSVGLGGKTAYDYNNLFPNKG YIGSVW 334
DB 233 -----EPLTG-----TG-----KIGGMLA----- 246
OY 335 WMDKRYETTLAAGISOPRNYRCNYWTSNYSNRSTTONLEKRAFSGGIWYVRDRAGIDAR 394
DB 247 -----LRGECTFPVQADFRSGNT-----RVAFDGVNDPMKMGVDLR 285
OY 395 LGAEPFLAERKIPGSDIDLGNHATMLTASMKRLLNNVLPENGHY--LDGKIGTTLGTF 453
DB 286 L-----KFSGD--SLGOLYB-----LTGVLPLPTPPFETDGR----- 316
OY 454 LSSTALIRTSARAGYFFETPENKKIGTFIIRGOAGYTVARD-----NADVPS----- 499
DB 317 ---VAKIDTEKSSVDFYRGFNIGRIGDSIDHGSILVYTTGKPRKLEGDVESROLRLADLGP 373
OY 500 --GLMFRSGGASSVRGYELDSIGLAGPNSVLPERALLVGSLEYQLPFTTLGSAVFHDM 557
DB 374 LIGVDSGKGAESKRSEKKGESVQPAKV-----LPYDRFETDK--WDV 417
OY 558 GDAANFKRMKLGKSGGLGVRFSP-----LAPFSFDIAYG-----HSDKK 598
DB 418 MDADVREKGRRIEHSGLSPISDLSTHILKLNADRLQLPLKFGMAGGSAIANHLEGDKK 476

RESULT 27
MRKC_KLEPN
ID MRKC_KLEPN STANDARD; PRT; 828 AA.
AC P21647;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Outer membrane usher protein mrkC precursor.
GN MRKC.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IA565;
RX MEDLINE=91100388; PubMed=1670938;
RA Allen B.L., Gerlach G.F., Clegg S.;
RT "Nucleotide sequence and functions of mrk determinants necessary for
RT expression of type 3 fimbriae in Klebsiella pneumoniae."
RL J. Bacteriol. 173:916-920(1991).
CC -!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE TYPE 3
CC FIMBRIAL SUBUNIT (MRKA).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL; M55912; AAA25095.1; -
PIR; D39142; D39142.
InterPro; IPR000015; Fimb_usher.
Pfam; PF00577; Usher; 1.
PROSITE; PS01151; FIMBRIAL_USHER; 1.
```

```

KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 828 OUTER MEMBRANE USHER PROTEIN MRKC.
FT DISULFID 813 827 POTENTIAL.
SQ SEQUENCE 828 AA; 91049 MW; B30EDF5798249FC9 CRC64;

Query Match
Best Local Similarity 19.8%; Score 110; DB 1; Length 828;
Matches 127; Conservative 84; Mismatches 213; Indels 216; Gaps 36;

QY 55 VRIDTQDSIKDWVHEHLPLTIQOQEEVLDKEQGFIA-----EAPDNVKTW- 102
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 VNIATDNNFPGDMRSYAP-----EIRDAVASNAUATVRQSNIIYQTTVPFGFTLQ 319
QY 103 -LRSKGYFSS-KVSLTEKDGAYTHITPGPRTKIANVGVAIGDILSDGNLAERYNA- 158
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 DVPYSGYSGDLEVSKEGSGVEFSVP-----YASVAQLLRPG-MRYALSACK 368
QY 159 -----LEN-----WQOPVGSDF-----DQSWENSKTSVLGAVTRKGYPLAKGNTRA 201
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 VDDSLRNKPMIYQGTWQHGLNNLFTYGTGTGDDYQAFLLGTGTMNTGIGALSFDVTHT 428
QY 202 AVNPDTATVDLNVVDSGRPAFGDFEITGTQRYPE---QIVSGLARFQPGTPYDLDLL 258
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 RLKSDT-----LDEHSG-----SYRATFNRMFETETQTSIVLAAYRYSTKGYNNLN- 473
QY 259 DFOQALBQNGHY-SCASVOADFRLQGRVPVKVSVTEVKRHKLETGIRLDSEYGLGGKI 317
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 DALYAVDQEKNTNSYTWRO-----KNGMTFTVYNQLPDGW---GGFYLCGRV 519
QY 318 AYDYN-----LFNKGYIGSVVMDM-----KYETTLAAGISQP----- 351
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 A-DYWNRSCTEKQYQFSYNNMY-CRLSWSVDAQRYVTPDSSGHRDRDRVSLNFSYPLWFG 577
QY 352 RNYRGHWTSNYSYRSTONLEKRAFGSGGIWYVRDAGIDARLGA- 402
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 ENRTANL-TSNTAFNNR-----PASSQIGVNGSLDSENNLNMGVSTTTAT 622
QY 403 GRKIPGSDIDLGNSHATMLTASWKRQLNNVLPENGYLDGKIGTGTGTFASSTALTRT 462
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 623 GRO---HDVALNGSVRTPTWTL-----NGSYSGEGYRQSGVGASGTLI--- 663
QY 463 SARAGYFTPEKNKLGTLFIRQAGYTVARDNADVPGLMFRSGGASSV----- 511
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 AHQGVGVFSPE-----TCPTMALIEAKDAAGVMLPGSPGTRIDSNGYAILPYL 711
QY 512 RGYELDSIGLAGPNS-----VLPERALLVGSLEYQLPFTRTLSCAVFHDMDG 559
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 RPYRINSVEI-DPKSGNDVAFGTVAGVVPWEGSVV-----KVSFDTTLQNNI- 759
QY 560 AANFRKMLKHGSLGVWRFSPPLAFSGFDIAYCHSDKKI 599
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 760 -----TLRARQANGL-----PL-PFAATI-FGPSKEI 785

RESULT 28
FRUA_STRMU STANDARD; PRT; 1423 AA.
AC Q03174;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fructan beta-fructosidase precursor (EC 3.2.1.80) (Exo-beta-D-
DE fructosidase) (Fructanase).
GN FRUA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCB1_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=93014178; PubMed=1398976;
```

```

QY 409 SDIDLGNSHATMLTASMKROLLNNVLPENGHVLDGKICTTLGTLFSLSTALIRTSARAGY 468
Db 1057 YDIDL-----KVQGLIN-----LFVASGNIDPSQAYSVQ 1086

QY 469 FTFPENKKLGTFTIRGOAGYTVARDN 494
Db 1087 FGDSETVRLYRF-----AGDTIAEAN 1107

RESULT 29
TACY_STRPY
ID TACY_STRPY STANDARD; PRT; 571 AA.
AC F21131;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Streptolysin O precursor (Thiol-activated cytolysin).
GN SLO OR SP70167.
OS Streptococcus pyogenes
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88057628; PubMed=3502717;
RA Kehoe M.A., Miller L., Walker J.A., Boulnois G.J.;
RT *Nucleotide sequence of the streptolysin O (SLO) gene: structural
RT homologies between SLO and other membrane-damaging, thiol-activated
RT toxins.;
RL Infect. Immun. 55:3228-3232(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szecate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT *Complete genome sequence of an M1 strain of Streptococcus pyogenes.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -!- FUNCTION: SULFHYDRYL-ACTIVATED TOXIN. IS ABLE TO LYSE CHOLESTEROL
CC CONTAINING MEMBRANES. CAN BE REVERSIBLY INACTIVATED BY OXIDATION.
CC CHOLESTEROL IS THE RECEPTOR FOR THE BINDING OF THESE TOXINS TO
CC EUKARYOTIC CELL MEMBRANES.
CC -!- SIMILARITY: BELONGS TO THE THIOI-ACTIVATED CYTOLYSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M18638; AAA26975.1; -.
CC EMBL; AE006485; AAK33267.1; -.
CC PIR; A43507; A43507.
CC HSSP; P19995; 1PFO.
CC InterPro: IPR001869; Thiol_cytolysin.
CC Pfam: PF01289; Thiol_cytolysin; 1.
CC PRINTS; PR01400; TACTIOLYSIN.
CC ProDom; PD007062; Thiol_cytolysin; 1.
CC PROSITE; PS00481; THIOI-CYTOLYSINS; 1.
CC Toxin; Hemolysin; Cytolysin; Signal; Complete proteome.
CC SIGNAL 1 33
FT CHAIN 34 571 STREPTOLYSIN O.
FT ACT_SITE 530 530 BINDING TO CHOLESTEROL (BY SIMILARITY).
SQ SEQUENCE 571 AA; 63638 MW; D05AA9979DCBACAB0 CRC64;

Query Match 3.48; Score 109; DB 1; Length 571;
Best Local Similarity 19.08; Pred. No. 9.4;
Matches 124; Conservative 92; Mismatches 234; Indels 202; Gaps 31;

```

```

QY 8 LLLPALFFPHAYAPADLSKNGAAGFALPKNKSPDSESVLKPKPVRIDTQDSEIKDM 67
Db 16 LUTAAIIGNLVTAENSKONTASTETTTTNEQPKPESSELTEKAGOKTDDMLNSNDM 75

QY 68 VE---EHLPLITQQQOEYL--DKBQTGLABEAPDNVKTMLRSKGYFSSKVSUTEKDAY 122
Db 76 IKLAPKEMPLESAEKEKKSEDKKKS--EDHTEELNDKIYSLNY--NELEVLAKNGET 130

QY 123 TVRITPGPRTKIANGVAILGDILSDGNLAFYRNALENWQOPVGSDFQDSDWENSKT-- 180
Db 131 IENFVPEKGVKKDKFIVI-----ERKKKNINTTP 160

QY 181 ---SVLGAVTRKGYPLA-KLGNTAAVNPDATVVDLVNVVDSGRPIAFGDFETGTQRY 236
Db 161 VDISITSDVTDRTPAALQALANKGFTENKPDVV-----TKRNP 199

QY 237 EGI-----VSLGARFOPGTPYDLDLLDFOQALEQ--NGHYSGASVQADPDRLQGRVPV 289
Db 200 QKIHDLPCMGDKATVEVNDPTYANVSTADNLVNHQHDNYSG-----GNTLPA 248

QY 290 KVSVTE---VKRHKLETGIRLDSEYGLGGKIAYDYINLFNKG-----YIGSV 333
Db 249 RTQYTESMVYSKSOIEAALNVNSKI-LDGTIGIDFKSI-SKGEKVMIAAYKQIFVTSA 306

QY 334 -----VWQMD-KYETTLAAGISQPNRYNGNYTNSVNSTTQNLKRAFSGGIYW 384
Db 307 NLPNPNADVDEKSVTFKELQKGVSV---NEAPPLFVSNVAGRTVFKLETSSKNDV-- 361

QY 385 VRDRAGIDARL-GAEFLAEGRKIPGSDIDLGNS-----HATMLTASMKRQLL 430
Db 362 ---EAFSAALKGTQDKTNGKY---SDI-LENSSTFATVVLGGDAAEHNVTKDF--DVI 412

QY 431 NNVLHPENGHYDGKIGITLGLTFLSSTALIRTSARAGYFTTPENKKLGIIRGOAGYTV 490
Db 413 RNVI-----KDNATFRKKNPAPYISYTSV-----FLKNKKTAG---VNNRTEY-V 453

QY 491 ARDNADVPSGLMFRSGGASSVRGYEL--DSIGLAGPNSVLPERALLVGSLEYQLPFTT 548
Db 454 ETTSTEYTSKGLNLSHOGAYVAQYELLWDEIN----- 485

QY 549 LSGAVFHDMDGDAANFKRMKMLKHGSLGVRWFSPFLAPFSDIAGVHSDKKIR 600
Db 486 -----YDDKGKEVITKRW-----DNNWYSKTSFSTVPLGANSRNR 524

RESULT 30
FHAC_BORPE
ID FHAC_BORPE STANDARD; PRT; 584 AA.
AC P35077;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hemolysin activator-like protein FHAC precursor.
GN FHAC.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RA Williams R.J.L.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO P. MIRABILIS HPMB AND S. MARCESCENS SHLB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64876; CAA46092.1; -.
CC DR

```

```
DR MEROPS; M19. UNW; -.
KW SIGNAL. 1 30 POTENTIAL.
FT CHAIN 31 584 HEMOLYSIN ACTIVATOR-LIKE PROTEIN FHAC.
FT CHAIN 31 584 HEMOLYSIN ACTIVATOR-LIKE PROTEIN FHAC.
SQ SEQUENCE 584 AA: 64445 MW: 8562352C07702178 CRC64;

Query Match 3.4%; Score 109; DB 1; Length 584;
Best Local Similarity 20.5%; Pred. No. 9.7;
Matches 128; Conservative 71; Mismatches 238; Indels 186; Gaps 28;

QY 22 PAADLS-ENKAGALFKNKSPTDSVSLKPKFPVRIDTQDSEIKDMVEHPLITQQQE 80
Db 64 PPVELNPQSEAAAPA-----RKPDATSGHTVTVHVDL---DFGVEGRLFPAPLVQDYLN 116
QY 81 EVIDKQGTGLAEAPDNKTKMLRSKGYFSSKSLTEK---DGAYTVHIIPGPRTKIANV 137
Db 117 RPLDNEQL-FLLVKA---LSAALYDRGYATSVTFVPPGVVDGVLKLVKVEWG-RIK--- 167
QY 138 GVALIGDILSDGNLAEYR-----NALENKQOPVGSDFDQDSEKTSVLCVATKGY 191
Db 168 -----GWLIDGKPLECTKDRMKVPSMPGMDKVLNVFDID----- 203
QY 192 PLAKLGNTRAAVNPDPATVLDLVNVVDSGRPIAFGDFEITCTQRYPEQIVSGLARFGPTP 251
Db 204 -----QAIYNINNGGTGNTITVPADYGYSLVDLQRRALPRVSLGMDNSGPGTP 255
QY 252 YLDLILLDFQAL-EQNGHYS-CASVOADFRLQGRVPKVSVEVRHKLETGIRLDS 310
Db 256 -----ENGRKYNAVSTAN-----DLGLNDTL-----GLYIGNR 285
QY 311 YGL-GKIAVDYVNLNKGIVGSVYVMDKYEYTTLAAGISQPRNY-----RGNYWT--SNV 363
Db 286 YYRDAGHDAERNYDLMSYVPLGRTRLD-----LOTGYSTYRNLTKTRYGOYSAGNSR 338
QY 364 YNRSTTONLEKRAFGSGGIWYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGSHATWLT 423
Db 339 SFGKATRLLYTRDSQFSVY---GGLKLRQNKNYLA-----GTRLDVSSKHYSDVTV 388
QY 424 SWKRQLNNVLPENGHYLD-----GKIGT-----TLGTFSLSTALITSA 464
Db 389 GMYSTQRG-----NAYFGDLSTFRGVGVNNGKYAAYDERGPGQNVSRFNSGLAWTRYMA 444
QY 465 RAGYFTPTNKKLGTFIIRGOAGYVARDNADVPGLMFRSGGASSVRYGEL-----DS 518
Db 445 LAQO-----PLOWASQLGFOYSRQ--QLNLSVOITVGDYTVRGVNLNLTSSGDS 492
QY 519 -----IGLAGPNSGVLPERALVGSLEYQLPFTRTLSGAVFHDMDGAAANFK 565
Db 493 GYVLSNTLTPVPQFSLGKQASVAPFVGADVGALKSNHPDARTIRMA----- 539
QY 566 RMLKKGSGLGVRWFSPAPPSF 588
Db 540 -----GLAAGVREFDLPYARMSF 556

RESULT 31
CARB_VIBCH
ID CARB_VIBCH STANDARD; PRT; 1076 AA.
AC Q9KPH9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
GN phosphate synthetase ammonia chain).
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Emolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -I- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O -> 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -I- COFACTOR: Binds three manganese ions (By similarity).
CC -I- PATHWAY: Arginine biosynthesis.
CC -I- PATHWAY: Pyrimidine biosynthesis; first step.
CC -I- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -I- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: A6004309; AAF95532.1; -.
CC HSP: P00968; IAGX.
CC TIGR: VC2389; -.
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_like.
DR Pfam: PF00289; CPase_L_chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CPASE.
DR PROSITE: PS00866; CPASE_1; 2.
DR PROSITE: PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 404 553 OLIGOMERIZATION DOMAIN.
FT DOMAIN 554 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 936 1076 ALLOSTERIC DOMAIN.
FT REPEAT 1 553
FT REPEAT 554 1076
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 303 354 ATP (POTENTIAL).
FT METAL 285 285 MANGANESE 1 (BY SIMILARITY).
FT METAL 299 299 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 2 (BY SIMILARITY).
FT METAL 828 828 MANGANESE 3 (BY SIMILARITY).
FT METAL 840 840 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1076 AA: 117911 MW: 8286623B247E4E1 CRC64;

Query Match 3.4%; Score 109; DB 1; Length 1076;
Best Local Similarity 20.8%; Pred. No. 23;
Matches 138; Conservative 80; Mismatches 260; Indels 184; Gaps 31;

QY 27 SENKAAGFA-----LFPKNKS-----POTESVKLKPKE---PV--RIDTQD 61
Db 474 AEVKAAGFAGLNDVLRKMKRKGFSARLSKLLGLVSENEIRRLRDYNIHPVKRVDTC 533
QY 62 SEIKDMVEHPLITQQQEVLQDQGTGLAEAP-DNVKTMLRSGK-----YF 109
Db 534 AEPKS-----DTAYMYSTYDEE-----CEANPTDKDKIMVLGGGPNRIGQGIEFDYC 580
QY 110 SSKVSLT-EKDGAYTVHIIPGPRTKIANVGA-----ILGDILS-----DGNLAE 153
```

Db 581 CVHAALALREDGYETIMVNCNPETVSTDYDTSRLRPEPVTLDVLAIVRVEPKGVIVQ 640
Qy 154 Y-----YRNALENWQVY-CSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAVNP 205
Db 641 YGGOTPLKLARALEAGVPICTSPDAIDRAEDRERFQQAVALGLK-----QP 689
Qy 206 DTATVD-LNVVVDGSRPIAFG-----DFEITGTQRYPEOIVSG-----242
Db 690 DNATVTAIEQATEKSREIGFPLVVRPSYVLGGRAMEIVYDEODLRRYFNEAVSVNESPV 749
Qy 243 -LARF-OPGTPYDLDLLD-----FQQALEQNGHYSGASQVADFDRLQGRVVPKVS 292
Db 750 LLDRFLDDATEVDVDAICDGERVVGIMHEIEQAGVHSGDSACS-----LPAYTL 800
Qy 293 VTEVKRHKLETGIRLDSYGLGKTIAYDYNLFNKGYIGSVWMDMKYETTLAAGISQRP 352
Db 801 SOEIODKMEQVEKLAFLGVLGRGLNIQFAVKDNEVYLIEV-----NPR 844
Qy 353 NYRGNWTS-----NYSYNRSTONLEKATSGGI-----WTVRRA¹-----GIDA 393
Db 845 AARTVPFVSKATGAPLAKTAARVMVGOTLEOQGTKEITIPPYSYKVEVLPFNKFFGVDP 904
Qy 394 RLGAELFLABGRKIPGSDIDLGNSHATMLTASMKRQLLNVLHPENGHYLDGKIGTTLGTF 453
Db 905 LLGPEMRSTG-----EVMCGVATFAEYAKAEKCGSVYPEGGRAL-----LSVREGDK 953
Qy 454 LSTALIRTSARAGYFFTPENKKGITFIIRGQAGTYVARONADVPSGLMFRSGASSVRG 513
Db 954 QRVVDLASKLVKGLQYLDATH---GTAVILGEAGINPLVN-KVHEG---RPHILDRIN 1006
Qy 514 YELDSI-GLAGPNSGYLPERALLVGSLEYQLPFTTISGAVFHDH---GDAAANFRMKL 569
Db 1007 HEVTYIVNTASGRQAEDESKVLRRLCALAHKVNYYITLNAAFATCMSHTADAKASVTSQVE 1066
Qy 570 KH 571
Db 1067 LH 1068

RESULT 32

MYFC_YEREN STANDARD; PRT; 841 AA.
AC P33408;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane usher protein myfc precursor.
GN MYFC.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=W1024 / Serotype O:9;
RC MEDLINE=94018646; PubMed=8105362;
RA Irlarte M., Vanooteghem J.-C., Delor I., Diaz R., Knutton S.,
RA Cornelis G.R.
RT "The Myf fibrillae of Yersinia enterocolitica".
RL Mol. Microbiol. 9:507-520(1993).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE MYFA
CC FIMBRIAL SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC EMBL; Z21953; CAA79953.1; -
DR PIR; S39365; S39365.
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 841 OUTER MEMBRANE USHER PROTEIN MYFC.
FT DISULFID 817 840 POTENTIAL.
SQ SEQUENCE 841 AA; 93414 MW; B6D207F75161D0D0 CRC64;

Query Match 3.4%; Score 108.5; DB 1; Length 841;
Best Local Similarity 18.7%; Pred. No. 17;
Matches 124; Conservative 82; Mismatches 232; Indels 225; Gaps 32;

Qy 16 FPHAYAPADLSENKAA---CFALFKNKSPTDE-----SVKLKPKFPVRIDTQ 60
Db 111 YPELLPPEKNIQDKQASQCVNLAIPAASEEFYAMRLVLNIPQVALPKDPEIPVERW 170
Qy 61 DSEIKDMVEEHLPIITQOOEEVLDKEQTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG 120
Db 171 DGGI-----TAFLLNYMANSTTTTYRQTG-----EQQS 198
Qy 121 AYVHIPTGPRKIANVGVAILGDILSDGNTAEYTRNALENNWQPVGSDTQDSWENSKT 180
Db 199 SHYVQLYPG-----FNIG-----AWRIRNA-TSMNQ---SGNSACKWQSSYI 236
Qy 181 SVLGAVTRKGYPLAKLGNTRAAVNDPTATVD-LNVVVDGSRPIAFGDFEITGTQRYPEOI 239
Db 237 ----YATRGLYRL---KSRVTLGOSYTPGDFDSVPFSGVYMLGDDNNMLPSSORDPIV 288
Qy 240 VSGLARFO-----PGTPYDL-DLLDFOQ-----ALEQNGHYSGAS 274
Db 289 VRGIARQARVEVRONGYLLYSTVVSPPG-PFELTDILPSHNSGDLHVTVLESNG-----341
Qy 275 VOADFDRLQGRVPRKVSYTEVRHKLETGI-----RLDSE-----YGLGK 316
Db 342 -----TTQOFTVPYSVPAIRLRLKGRLYNLNLMAGRYRPNVDVETTTPIAQTAVAYGLP--393
Qy 317 IAYDYNLFNFKGYIGSVVWMDKYEYTLAAGISQRPNGYTRNGYNTSNYSNSTQNLEKR 376
Db 394 -----WNL--TAEVGG-QMSPHYQATTAGMV-----MLGQYGLASSITQATSEYKQOQ 440
Qy 377 AFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASMKRQLLNVLHP 436
Db 441 PVKGQVWEVR-----VNKTQASDTSFVSVNSQYSTADFSN--LSDVLQS 483
Qy 437 ENGHYLDGKIGTIGTTLGTFLSSTALIRTSARAGYFFTPENKKGITFIIRGQAGTYVARONAD 496
Db 484 YRRHDYSRR-----DWHNSLRNTHYVVG-----OPLGQF---GYLNLNWSRON--525
Qy 497 VPSGLMFRSGGASVRGYELDSIAGLPNGSVLPERALLVGSLEYQLPFT-----R 547
Db 526 -----YRDAPASSSMGVQY-----SPNIGNLVCSLDWTQNOYRGNQDR 563
Qy 548 TLSGAVPHDGDAAANFRMKLKHGSLGCLVGRWFSPLAFSFDIAYGHS-DKKIRWHISLG 606
Db 564 LLSLSVMPILGRERDITYAAYRMTSSDNSKDHMS-----LYGHAFDRNLUSMNVROT 614
Qy 607 TRF 609
Db 615 ERY 617

RESULT 33

GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (Dextrantransferase) (Sucrose 6-glucosyltransferase).
GN GTF-C.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RL "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5";
RT Gene 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + ((1,6)-alpha-D-glucosyl)(N) = D-
CC fructose + ((1,6)-alpha-D-glucosyl)(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isp-sib.ch).
CC -----
DR EMBL: M22054; AAA88592.1; -;
DR EMBL: M17361; AAA88589.1; -;
DR PIR: J03435; J03445.
DR PIR: C33135; C33135.
DR InterPro: IPR002479; CW_binding.
DR InterPro: IPR003318; GH_70.
DR Pfam: PF01473; CW_binding_1; 7.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; repeat; Dental caries.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
FT REPEAT 1375 1375 A REPEAT (INCOMPLETE).
FT SIGNAL 1375 1375 AA: 153022 MW: 44880CBEEDAAAE13 CRC64;
SQ
Query Match 3.4%; Score 108.5; DB 1; Length 1375;
Best Local Similarity 19.7%; Pred. No. 35;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;
QY 30 KAAGFALPKNSPTESVCLKPRFPVRIDTQDSEIKDMVEEH-----LPLTQOQEEV-- 82
DB 729 RTSGVAVIEGNPS-----LRLKASDRVVVMGAHKNQAYRPLLTDTGDIKA 777

QY 83 --LDKEQTGEL--AEEAPDNVKTMLRSKGYESSKVS-----LTKDGAYTVIHT 127
DB 778 YHSDQEAAGLVRYTNDRGELIFTAADIKGTANPQVSGYLGWVPVGAADQDVRVAASTA 837
QY 128 PGPRTKIANGVAITLSDILSDG-----NLAEYRNA-----LENMQQVPGSDFDQDS 174
DB 838 PSTDGKSVHQNAALDSRVMEPGFSNFQATKKEEYTNVVTAKNVDFKAENGVTDFMAP 897
QY 175 WENSKT--SVLGAVTRKGYPLAKLNGNTRAANVPDPTATVDLNVVDSGRPIAFG----- 225
DB 898 QYVSTDCGSLDSVIQNGYAF-----DTRVDLGI-----SAPNKYGTADDLVK 940
QY 226 -----DFEITGTRYPE-QIVSGLARFPQGP-----YDLDLLDF 260
DB 941 AIKALHSKGIKVMADWVPDQYALPEKEVVTATRVDKYCTPVAGSQIKNTLYVVD---GK 997
QY 261 QOALEQNHYSGA---SVQADFRLQGRVPKVSVTEVKRHKLFTGLRLDSEYGLGKI 317
DB 998 SSGDQQAQKYGGAFLLELQAKYPEL-----PARKQISTGVPMPDPSVKIKQWS 1044
QY 318 A--YDYNLNFNKGYSVWDM--DKYETTLAAGISQPRN-YRGNWTSN----- 362
DB 1045 AKYFNGTNILGRG-AGYVLKDOATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVFDGK 1103
QY 363 --VSVNRSTTONLEKRAF--SGGIWVVRD-----AGIDARLGAE--FLAEGRKIPGSDI 411
DB 1104 GYVY--STSCNQAKNAFISLGNWYFNDNGYVWVTGAOSINGANYFSLNSGIGLRNAYI 1161
QY 412 DLGNHATMLTASWKROLLNNVLPENGHYLGCKTGLTGLTFLSTALIRTSARAGYFT 471
DB 1162 DNGKNVLSYTGNDGRY-----ENGYLFQQQWRYPQNGIMAVGLTRVHGAVQYFDA 1213
QY 472 PENKLGITFIIR--GOAGYTVARNADVPGLMFRSGASSVRGYELDSIGLA----- 522
DB 1214 SGFOAKGOFITTAGDKLRY-FDRSGNQISNRFVRN---SKGENFLDHNGVAVTGVTF 1269
QY 523 -----GPNQSVLPERALLV-----GSLEYQLP-----FTTSLGAVF-HMDGAA 561
DB 1270 NGORLYFKPNG--VQAQGEFIRDANGLYRYDFNSGNEVNRNFRVRSKGEWFLFDHNGIA 1327
QY 562 ANFKRMKLKHGSGGLGVRF 580
DB 1328 VTGARVVVNGHASILLSLVF 1346

RESULT 34
HSF2_MOUSE
ID HSF2_MOUSE STANDARD; PRT; 517 AA.
AC P38533;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heat shock factor protein 2 (HSF 2) (Heat shock transcription factor
DE 2) (HSTF 2).
GN HSF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=92009180; PubMed=1717345;
RA Sarge K.D., Zimarino V., Holm K., Wu C., Morimoto R.I.;
RT "Cloning and characterization of two mouse heat shock factors with
RT distinct inducible and constitutive DNA-binding ability";
RL Genes Dev. 5:1902-1911(1991).
CC -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION IN HIGHER
CC EUKARYOTES. HSF IS UNABLE TO BIND TO THE HSE UNLESS THE CELLS ARE
CC HEAT SHOCKED. HSF2 IS EXPRESSED IN A FORM THAT BINDS DNA
CC CONSTITUTIVELY BUT LOSES DNA BINDING BY INCUBATION AT GREATER THAN
CC 41 DEGREES C.


```

Db 277 ---INSVKDTTGVESIDENGKLLTSTRGRIKIEGNIGR--GAFINPNWLENYGRISL 331
QY 460 IRTSARAGYFFTENKK---LGTFTIRQAGTYVARD-----NADVPGLMFRS----- 505
Db 332 VKNDGK-DLLISTNLSAIGFCGTGNNISVLSRESKGQIDANVADMGFNANKGNIL 390
QY 506 GGASVRYGELDSIGLAPNGSVLPERALLVGS-LEYQLPFTRLS-----GAVPH-- 555
Db 391 GGYSSVSAY-MSSTGSGFSGS-----GFSVSGKNYSTGFANTIAISAQSLSAVYVNS 444
QY 556 -----DMGDAANFKRMKLKHSGGLQVR 578
Db 445 AGSGFSSGNSLQFATMK7SAGNTLGVK 472

RESULT 36
PRTC_ERWCH
ID PRTC_ERWCH STANDARD; PRT: 479 AA.
AC P16317;1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Secreted protease C precursor (EC 3.4.24.-) (Proc).
GN Erwinia chrysanthemi.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=B374;
RX MEDLINE=91009140; PubMed=2211614;
RA Deleplaire P., Wandersman C.;
RT "Protein secretion in Gram-negative bacteria. The extracellular
RT metalloprotease B from Erwinia chrysanthemi contains a C-terminal
RT secretion signal analogous to that of Escherichia coli alpha-
RT hemolysin."
RL J. Biol. Chem. 265:17118-17125(1990).
RN [2]
RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=B374;
RX MEDLINE=89255387; PubMed=2722818;
RA Deleplaire P., Wandersman C.;
RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are
RT synthesized and secreted as zymogens without a signal peptide."
RL J. Biol. Chem. 264:9083-9089(1989).
CC -1- COFACTOR: THE DIVALENT CATIONS MG(2+) AND CA(2+) INCREASE
CC PROTEASE ACTIVITY SIGNIFICANTLY ABOVE THE CONTROL VALUE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR
CC SECTION OF THIS METALLOPROTEASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59229; AAA24860.1; -
DR EMBL; J04736; AAA44862.1; -
DR EMBL; M60395; AAA63638.1; -
DR PIR; A38307; A38307.
DR HSP; P23694; 1AFO.
DR MEROPS; M10.054; -.
DR InterPro; IPR001343; Hemolysin_Ca_bind.
DR InterPro; IPR000130; Zn_Metpeptidase.
DR Pfam; PF00353; hemolysin_Catbind. 3.
DR PRINTS; PR00313; CABNDNGRPT.
DR SMART; SM00235; ZnMc; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.

```

```

DR PROSITE; PS00330; HEMOLYSIN_CALCITUM; 1.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Repeat.
FT PROPEP 1 17
FT CHAIN 18 479 SECRETED PROTEASE C.
FT METAL 188 188 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 189 189 BY SIMILARITY.
FT METAL 192 192 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 345 386 5 X REPEATS, GLY-RICH.
FT REPEAT 345 350 1.
FT REPEAT 354 359 2.
FT REPEAT 363 368 3.
FT REPEAT 372 377 4.
FT REPEAT 381 386 5.
SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;

Query Match 3.4%; Score 107.5; DB 1; Length 479;
Best Local Similarity 22.4%; Pred. No. 9.1; Mismatches 168; Indels 135; Gaps 27;
Matches 104; Conservative 58;

QY 88 TGFLAEAPDNVKTMLRSKYFS-SKVSLETKDGYVTHITPGPRTKIANVGVAILGDIL 146
Db 92 TGFVKFAEQIEQAKLSQSWSDVANLTFTEVTGNKSANITFGNTRDAS----- 141
QY 147 SDGNL-----AEY---YRNALENWQOPVGSDFDDQDSWENSKTSVLGAVTRKGYPLAKL 196
Db 142 --GNLDYGTQAYAYPCNYQGAGSSWY-----NYQSNIRNPGSEEGYGRQTF----THEI 190
QY 197 GNTRAAVNPDATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFOPGTPYDLDL 256
Db 191 GHALGLAHPG-----EYN--AGEGDP--SYND-AVVAEDSYQFSIMSYWGENETGADY---- 238
QY 257 LLDFOQALEONGHYSGASVQAD---FDRLQDGRVPVKVSVTEVRKHLETLGIRLDSYGL 313
Db 239 -----NGHYGAPMIDDIAAIQRLYGANWTR-----TG---DSVYGF 273
QY 314 GGTAYDYNL--FNKGYIGSVVMDKYEYTTLAAGISQPRNYRGNYWTNSVSNRSTTQ 371
Db 274 NSNTRDFYATDSSKALIFS-VWD-----AGGDTDFDSG-----YSNNQRI 315
QY 372 NLEKRAFSGGIWYVRDRAGIDARLGAFLAEGRKI-----PGSDIDLGNSHATMLTASW 425
Db 316 NLNKGFS-----DVGGLGKNSV---IAHGVTIENAIGSGNDLIVGNSADNLOGG-- 364
QY 426 KRQLLNVLHPENG-HYLDGKIGTGLTFLSTALIRTSARAGYFFTPENKK-----LG 478
Db 365 ---AGNDVLVYGGAGADTLYG--GAGRDYFVYVSG--QDSTVAAYDWIADFQKIDKIDLS 417
QY 479 TFIIRGQAGYT-----VARDNADVPGLMFRSGCASSV 511
Db 418 AFRNEGQLSFVQDQFTGKGVEYMLQWDAANSITNLMLHEAGHSSV 462

RESULT 37
P1CP_PSESR STANDARD; PRT: 587 AA.
ID P1CP_PSESR
AC P42790;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pseudomonaspepsin precursor (EC 3.4.21.100) (Pepstatin-insensitive
DE carboxyl proteinase).
GN PCP.
OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33067;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 216-224.
RX MEDLINE=95014496; PubMed=7929375;
RA Oda K., Takahashi T., Tokuda Y., Shlbano Y., Takahashi S.;
RT "Cloning, nucleotide sequence, and expression of an isovaleryl
RT pepstatin-insensitive carboxyl proteinase gene from Pseudomonas sp.
RT 101."

```


RL J. Biol. Chem. 269:26518-26524 (1994).
 RN [2]
 RP SEQUENCE OF 216-585.
 RX MEDLINE=96157710; PubMed=8576087;
 RA Hayashi K., Izu H., Oda K., Fukuhara K.-I., Matsuo M., Takano R.,
 RT Hara S.;
 RT "The primary structure of pepstatin-insensitive carboxyl proteinase
 produced by *Pseudomonas* sp. No. 101.";
 RL J. Biochem. 118:738-744 (1995).
 RN [3]
 RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
 RP SITES.
 RX MEDLINE=99419069; PubMed=10488127;
 RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;
 RT Identification of catalytic residues of pepstatin-insensitive
 RT carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
 RL J. Biol. Chem. 274:27815-27822 (1999).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the B chain of insulin at 13-
 CC Glu-1-Ala-14, 15-Leu-1-Tyr-16 and 25-Phe-1-Tyr-26 and angiotensin
 CC I at 4-Tyr-1-Ile-5. A good synthetic substrate is Lys-Pro-Ile-Glu-
 CC Phe-1-Phe(NO2)-Arg-Leu.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- PTM: AUTOCATALYTICALLY PROCESSED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D37970; BAA07188.1; -
 DR MEROPS: S53.001; -
 KW Hydrolase; Protease; Serine protease; Zymogen; Periplasmic; Signal.
 FT SIGNAL 1 POTENTIAL.
 FT PROPEP ? 215 REMOVED IN MATURE FORM.
 FT CHAIN 216 585 PSEUDOMONAPEPSIN.
 FT PROPEP 586 587 REMOVED IN MATURE FORM.
 FT ACT_SITE 385 385 PROBABLE.
 FT ACT_SITE 502 502 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 543 543 PROBABLE.
 FT DISULFID 352 391
 SQ SEQUENCE 587 AA; 61072 MW; E193D8B2C225829A CRO64;
 Query Match 3.4%; Score 107; DB 1; Length 587;
 Best Local Similarity 20.4%; Pred. No. 13;
 Matches 137; Conservative 69; Mismatches 226; Indels 238; Gaps 34;
 QY 21 APAADLSENKAAGFALPKNKSPTDESVKLPKFPVTRDTQDSSEIKDWVEHLPLITQQQE 80
 DB 60 APLELAAGETAHIV-----SLKLR-----DEALKOLAQAVNPGNAQFG 101
 QY 81 EYLDKEQTQGLAEAPDNVNTMLRSKGYFSKVSLETKDGAYTVHTPGPRTKIANVGVA 140
 DB 102 KFLKRQ-----FLSQFA-----TEAQVQAVVAHLRKNFVNIHVPN-RLLSADGSA 150
 QY 141 ILGDILSDGN--LAEYVRNLENMCOOPVGSDFDSDSENSKTSVLGA--VTRKGYPLAKL 196
 DB 151 --GAVAAENTFLVYQLNGRAGYANTAPAOVDLGE-IVGSVGLQNVTR-AHPMLKV 206
 QY 197 GNTRAAVNPDTAT-----VDLNVVVD--SGRPIA-----FGDFEITGTQRYPEQIVSGLAR 245
 DB 207 GERSAAKTLAAGTAKGHNPTEPTTYDASSAPTAANTVGTITGGVSQ----- 255
 QY 246 FQPGTPTDLDLLDQQALEONGHVSAGASVQADFRLQGDVPVKVSVTEYKHKHLETI 305
 DB 256 -----TIQDLQQTTSNG-----LASVNTQTITG- 280
 QY 306 RLDSEYGLGKGIAYDYNLFNKNKGYIGSVVWDMKYEYTTLAAG--ISQPRNRYGNWYTS-- 361
 DB 281 ---SSNG-----DY-----SDQDQOGCEWDLDSQSIVSGAGGAVQQLLFYMQDQASGN 326

QY 362 ---NVSYNRSTTONLEKRAFGSGGWYVRDRAGIDARLGAE-----FLAEGRKIPGSDID 412
 DB 327 TGLTOAFNOAVSDNYAK-VINVSLGWCEDANADGTLOAEDRIFATAAAGGQTFSSVSGD 385
 QY 413 LGNSHATMLTASWKRQLLNVLHPENGHYLDGKI-----GTTLTGTFLLSS 456
 DB 386 EG-----VYECNNRCYPDGSTYSYSPASSPNVIAVGTTLYT----- 423
 QY 457 TALIRTSARAGYFFTPENKKGKLTGFIIRGOAGYTVARONADVPGLMFRSGGASSVRGYE- 515
 DB 424 ----TSAGA-----YSNETVMNEGLD-SNCKLWATGGGYSV--YES 457
 QY 516 ----LDSIGLAGPNSVLP-----ERALLVGSLEYQLPRTFLSG----- 551
 DB 458 KPSQSWVSQSGTGERLLPDISFDAAGTGALINYVGOIQITGGTSLASPIFVGLMARLOS 517
 QY 552 AVFHDMDGDAANF-----KRMKLKH-----GSLGCVRWFSPLAPFSFDIA--- 591
 DB 518 ANSNISLGFPAASFYSAISSTPSLVHDVKSGNNGYGGYNAGTGMWYPTGWSGLDIKLUS 577
 QY 592 -----YGH 594
 DB 578 AVIRSNFGCH 587
 RESULT 38
 K6PP-RABIT
 ID K6PP-RABIT STANDARD; PRT; 791 AA.
 AC P47859;
 DT 01-FEB-1996 (Rel. 33; Created)
 DT 01-FEB-1996 (Rel. 33; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE 6-phosphofructokinase, type C (EC 2.7.1.11) (phosphofructokinase
 DE 1) (phosphohexokinase) (Phosphofructo-1-kinase isozyme C) (PFK-C).
 GN PFKP.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=94164929; PubMed=8119919;
 RA Li Y., Valaitis A.P., Latschew S.P., Kwiatkowska D., Tripathi R.L.,
 RA Campbell M.C., Kemp R.G.;
 RT "Structure and expression of the cDNA for the C isozyme of
 RT phosphofructo-1-kinase from rabbit brain.";
 RL J. Biol. Chem. 269:5781-5787 (1994).
 CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
 CC fructose 1,6-bisphosphate.
 CC -1- COFACTOR: MAGNESIUM.
 CC -1- ENZYME REGULATION: ALLOSTERIC ENZYME ACTIVATED BY ADP, AMP, OR
 CC FRUCTOSE BISPHOSPHATE AND INHIBITED BY ATP OR CITRATE.
 CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U01154; AAA17707.1; -
 DR HSSP: P00512; 3PEK.
 DR InterPro: IPR00023; Ppfuckinase.
 DR Pfam: PF00365; PFK; 2.
 DR PRINTS: PR00476; PFERCKINASE.
 DR PRODOM: PD000707; Ppfuckinase; 2.
 DR PROSITE: PS00433; PHOSPHOFRUCTOKINASE; 2.

KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme;

KW Phosphorylation; Magnesium; Multigene family.

SQ SEQUENCE 791 AA; 86349 MW; 3C10A36F229FD8E8 CRC64;

Query Match

Best Local Similarity 3.4%; Score 107; DB 1; Length 791;

Matches 118; Conservative 77; Mismatches 265; Indels 164; Gaps 25;

```
Qy 81 EYLDKEQTGLAEAPDNVKTMLRSKGFSSKVSUTEKDGAYTVHTPGP----- 130
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 EVIDAINTTAQSHORTFVLEVGRCGCLYALVSALA--CGADWFLPESPPEGWEQMC 253
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 131 -----RTKIANYGVAILDGLSDGNLAEEYRNALNNWQPVGSDFDQSWENSKTSVLG 184
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 VKLSENRAQKKRLNIIVAE-----GADITLNRPTS-----EKIKELV-- 292
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 185 AVTRKGYPLAKLNTAAVNPDTATVDLNNVVDSCRPIAFGDFELTGQRYPEQIYSLGA 244
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 -VTOLGY-----DTRVTILGHVORGTPSA---FORILASRMGVAVLALL 334
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 245 RFQPGTP-----YDLDLULD---FQQALEQNGH-----Y 270
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 EATPETPACVSVLSGNHVRVLPVECVQTOEQVKAMDERRRFDKDAVQLGRSPENLNTY 394
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 271 SGASVQADFRLQGRVPPKVSVTEVKRHKLETGIRLSEYGL--GGKI--AYDYNLPN 326
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 KRLAIKLPDDKIQKNCNVAVINVGAPAGMNAVRVVGIDGKHFAYVDGDFDGA 454
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 327 KGYISVYV--DMDKYETTLAAGISOPRNYRNYTSNYSNRSTQHLKRAFSGGIWY 385
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 KGQIKETRWGDVGWGTGGGSLTGKRIPLPGKYLEIATQTR--THNINAILIG----- 507
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 386 RDRADIDARLGAFLAEGRK-----IPGSDIDLGNSHATMLTASWK 426
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 ----GPAEYLGLELSAAREKHEEFCVPMVVPATVSNVPGSOFSGADTALMTITDTC 563
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 427 ROLLNNVLPENGHYLDGKIGTGLTFLSSALIRTSARAGYFTTPENKLGTFIIR--- 483
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 564 DRIKQASGTRKRRVFIETMGGYCG--YLANMGLAAGADAAYIFEE-----PDIRDLQ 616
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 484 GQAGYTVARDNADVPGLMFRSGGASSVRGYELDSI--GLAGPNG--SVLPERALLVGSLEY 541
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 617 SNVEHLEKMKTTIQGLVLRNESC--ENYTTDFIYLYSEERGVDCKKNVLGHMQQ 674
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 542 ---OLPRTTSLGAVFDMCDAAANKRMLKHGSGGLGRV-----FS 581
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 GGAPSPDRNFCTKI-----SARAMQWITTKLSPGKGRFVSDDSIKRNVLFG 730
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 582 PLAPFSFDIAYGSHDKKIRWHISL 605
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 PVAEKNETDFEIRIPKEQWLKL 754
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 39

AFAC_ECOLI

ID AFAC_ECOLI STANDARD; PRT: 859 AA.

AC P5317;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane usher protein afac precursor.

GN AFAC.

OS Escherichia coli.

OC Plasmid pili055.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A30;

RX MEDLINE=95095929; PubMed=8002584;

RA Garcia M.-I., Labigne A., le Bouguenec C.L.;

RT "Nucleotide sequence of the afimbrial-adhesin-encoding afa-3 gene

```
RT Cluster and its translocation via flanking IS1 insertion sequences.*;
RL J. Bacteriol. 176:7601-7613(1994).
CC -I- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF AFA-III AFIMBRIAL
CC ADHESIN SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -I- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -I- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: X76688; CAAS4117.1; -.
CC InterPro: IPR000015; Fimb_usher.
CC Pfam: PF00577; Usher; 1.
CC PROSITE: PS01151; FIMBRIAL_USHER; 1.
CC KW Outer membrane; Transmembrane; Fimbria; Transport; Signal; Plasmid.
CC FT SIGNAL 1 28 POTENTIAL.
CC FT CHAIN 29 859 OUTER MEMBRANE USHER PROTEIN AFAC.
CC SQ SEQUENCE 859 AA; 93805 MW; 1FF7FE3715EEFB93 CRC64;
CC
CC Query Match 3.4%; Score 107; DB 1; Length 859;
CC Best Local Similarity 20.1%; Pred. No. 22;
CC Matches 135; Conservative 79; Mismatches 189; Indels 270; Gaps 38;
CC
CC 57 IDTQSEIKDMVEEHLPLI-----TQOQEVLDKEQTGLAEAPDNVKTMLRSKGFYS- 110
CC 278 MSSDESVPVNLREFAVPVVGARTQARIEV---RQNGYL-----IQSTVAPGAFAL 327
CC 111 -----SKVSLTEKDGAYTVHTPGPRTKIANYGVAILDGLSDGL-----AEY 154
CC 328 TDLPTVGTSGSDIAQVTVLESDDGTAQVFTVPTTPAIA-----LREGYLKYNVTAGQ 377
CC 155 YRNALENWQPVGSDPDQDSWENSKTSVLGAVTRK--GYP--LAKLGNTRAAVNDPTATVD 211
CC 378 YRSS-----DDAVEHTS---LGQVTANYGLPWGLTVYGLQAGADDYQSAALG 421
CC 212 LN-----VVVDS-----GRPIAFGD-----FEITGQRYPEQIYSGLARF 246
CC 422 LGWSLGRGAVSLDTTHSRQOQKGDHYETGDTWIRYKNKSELTGT----- 467
CC 247 QPGTPYDLDDLDFQQALEQ---NGHYSASVQADFRLQGRVPPKVSVTEVKRHKLET 303
CC 468 -----SFTAASYCYSSDGYHTLPDV---LDTWRDDRYAVRHTENRSRR---T 508
CC 304 GIRLSEYGLGGKIAY-----DYNNLPNKGYIG---SVVWMDKYETTLAANGISQPNYR 355
CC 509 TLSLSQSLGQMGYVGLNGSRDEYRDRPHRDYFCASYSYSWN-----NLSLVNWSNRNRS 564
CC 356 GNY-----WTSNVSYNRSTTQNLK-----RAFSGG 381
CC 565 GYGGWRSRTEDSVSMMSVPLGRWFGGADNDISTTAQMRSTGQDTRYEAGLNGRAFDRR 624
CC 382 I-WYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHA--TWLTASW----- 425
CC 625 LYWDVRE-----QWVPGE-----SHADTSLRLNLTWYGTGTGMYYS 663
CC 426 ---KROLL-----NNVLHPENGHYLDGKIGTTL-----GTFLSALIRTSAR--- 465
CC 664 SSTMRQLNAGMSGSMVAHSE-GVTFQGTGDTVALTAAPGVSGASVGWGPVTRDFRGYT 722
CC 466 -AGYFFTPENKLGTFIIRCOAGYTVARDNADVP-----SCLMFRSGASSSVGYE 515
CC 723 LAGY-ASPYENVTLL-----DPTTFPEDAEVPQDTSRVVPTGAVVVRAGFRTRVGGRA 775
CC 516 LDSIGLAGPNGSVLP-----ERALLVGSLEYQLPPTRTLGSAGVFDHMDGAAAFKR 566
CC 776 L--VSLARODCTPLPFGAVVTVVEGERGAAG-----SAGVVVGDGRGYLS--- 818
```

Db 265 AVLAR-EHGIPFAVAPLSSIDMSLKSGKEIPIERSPEVLTCGGRIAPD----- 315

Qy 287 VPVKVSVTEVKRHKHLETGIRLD 308
Db 316 VPVYNPAFDVTPHKYVTGIITD 337

Search completed: November 9, 2002, 01:19:27
Job time : 30 secs

QY 567 MKLKHGSGLGVRW 579
Db 819 -GUKESGKLKAOW 830

RESULT 40
E2BI_PYRAB
ID E2BI_PYRAB STANDARD; PRT; 356 AA.
AC OS0216;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative translation initiation factor eIF-2B subunit 1 (eIF-2B
DE GDP-GTP exchange factor).
GN AIF-2B1 OR PAB2444.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_taxid=29292;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE EXCHANGE OF INITIATION FACTOR 2-BOUND GDP
CC FOR GTP (BY SIMILARITY).
CC -!- SUBUNIT: COMPLEX OF TWO DIFFERENT SUBUNITS (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE EIF-2B ALPHA/BETA/DELTA SUBUNITS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ248287; CAB50246.1;
CC InterPro: IPR005251; AIF-2BI_fam.
CC TrEMBL: P0000649; IF-2B.
CC Pfam: PF01008; IF-2B; 1.
CC TIGRFAMs: TIGR00512; aIF-2BI_fam; 1.
CC TIGRFAMs: TIGR00524; eIF-2B_rel; 1.
CC Hypothetical protein; Initiation factor; Protein biosynthesis;
CC Complete proteome.
SQ SEQUENCE 356 AA; 39548 MW; 4273C0A06C5B7513 CRC64;

Query Match 3.4%; Score 106.5; DB 1; Length 356;
Best Local Similarity 23.6%; Pred. No. 7;
Matches 76; Conservative 49; Mismatches 120; Indels 77; Gaps 17;

QY 21 APAADLSENKAAGFALF-----KNKSPDT-----ESVKLKPKFPVRIDTQDSEIKDM 67
Db 59 APA--ICAAAFGLALYAETSAKSKDEFMDGFKAYETLKNTRPTAVNLFNALNRKKL 116
QY 68 VEEHLPLITQQQEEVLDKQGTGLAE--APDNVYMLRSKGYFSSKVSLETKDGAYTV 124
Db 117 VEEHL-----EDPLDEIKSLIVNEAKIADVEDVEANLR-MHYGAELV---PEGNLLT 165
QY 125 HITPGPRTKIANGVAILGDIL----SDGNLAEYRYRNALENWQOPVGSDFDQDQSWENSKT 180
Db 166 HCNAG---SLATVHLTGAVYRVHVKDGLSKLLWDE----TRPVLGARLSAWEYS-- 216
QY 181 SVLGAVTRKGYPLAKIGNTRAANVPDPTATVDLNVVDSGRPIAFGDFEITGTORYPQIV 240
Db 217 -----YDGLANVKKLIADNAAAFVQQGLVDA-IIVGADRIVANGDF----ANKIGTYML 264
QY 241 SGLARFQPGTPY-----DLDLLDFQQALEONG-----HYSGASVQADFRLQGDOR 286

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 9, 2002, 01:23:59 ; Search time 1941 Seconds
(without alignments)
5081.428 Million cell updates/sec

Title: US-09-857-669-2

Perfect score: 3173

Sequence: 1 MHKPTALLPALFFPHAY.....IAYGHSKIRWHISIGTRF 609

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q/cy2_1/USPTO.spool/US09857669/runat_05112002_110008_7737/app_query.fasta_1.775
-DB-EST -qfmt=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09857669.ecgn_1.1045 -runat_05112002_110008_7737 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	141.5	4.5	792	14	B0506223 EST613638
C 2	129	4.1	607	10	AV557254 AV557254
C 3	125	3.9	649	17	BH366320 AG-ND-161
C 4	121	3.8	503	10	BE459389 EST414681
C 5	119	3.8	494	10	AW191135 687022D06
C 6	119	3.8	2776	11	AK002711 Mus muscu
C 7	117	3.7	1829	11	AV109520 Zea mays
C 8	111	3.5	588	10	AW053211 614073B08
C 9	109.5	3.5	503	11	AY108685 Zea mays
C 10	108	3.4	3045	11	AK004789 Mus muscu
C 11	107	3.4	1077	13	B1559235 603240861
C 12	105.5	3.3	353	14	D46516 RICS11241A
C 13	103.5	3.3	620	12	BF942509 1D2 CDNA
C 14	103.5	3.3	928	14	BQ723872 AGENCOURT
C 15	103	3.2	423	9	AT363592 SMOVL2CAS
C 16	103	3.2	2936	11	AK011769 Mus muscu
C 17	101.5	3.2	451	13	BM438066 VYAO30A02
C 18	101.5	3.2	1031	14	BQ876247 AGENCOURT
C 19	101.5	3.2	3095	17	BH770975 LLMGTAG70
C 20	100	3.2	724	14	BQ872838 OG116M16
C 21	99	3.1	613	14	BQ871231 OG111D13
C 22	99	3.1	1171	13	BM560625 AGENCOURT
C 23	98	3.1	622	14	BQ875918 OG1922.Y
C 24	98	3.1	994	12	BF535617 602054012
C 25	97.5	3.1	4944	11	BC035446 Homo sapi
C 26	97	3.1	630	14	BQ871709 OG112J17
C 27	96.5	3.0	590	14	BQ479447 Ku36h01.Y
C 28	96	3.0	577	14	BQ872562 OG116A10
C 29	96	3.0	660	14	BQ875018 OG16020.Y
C 30	96	3.0	677	14	BQ875030 OG16P11.Y
C 31	96	3.0	731	14	BQ874536 OG15M18.Y
C 32	96	3.0	735	14	BQ850925 OG13024.Y
C 33	96	3.0	757	14	BQ871570 OG112D16
C 34	96	3.0	1072	14	BM519327 AGENCOURT
C 35	95.5	3.0	563	14	BQ675465 AGENCOURT
C 36	95.5	3.0	1086	12	BG704719 602688326
C 37	95.5	3.0	2466	11	BC013324 Homo sapi
C 38	95	3.0	578	14	BQ875852 OG19E21.Y
C 39	95	3.0	581	14	BQ875483 OG18E01.Y
C 40	95	3.0	622	14	BQ874935 OG16K24.Y
C 41	95	3.0	659	14	BQ858721 OG11A18
C 42	95	3.0	660	14	BQ874292 OG14O19.Y
C 43	95	3.0	662	14	BQ875986 OG19K20.Y
C 44	95	3.0	669	14	BQ874486 OG15H01.Y
C 45	95	3.0	671	14	BQ875563 OG18H13.Y

ALIGNMENTS

RESULT 1
B0506223/c
LOCUS
DEFINITION B0506223 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STMGI92 3', end, mRNA sequence.
ACCESSION B0506223
VERSION B0506223.2
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 792)

AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.
JOURNAL Generation of a set of potato cDNA clones for microarray analyses
COMMENT Unpublished (2002)
 On Jun 10, 2002 this sequence version replaced gi:21365092.
 Other_ESTs: EST613637
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato@igr.org
 This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@rsgen.com
 Seq primer: T7.
 Location/Qualifiers

FEATURES

source

1..792
 /organism="Solanum tuberosum"
 /cultivar="Kennebec or Binjete"
 /db_xref="taxon:4113"
 /clone="STMG192"
 /clone_lib="Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes tubers, or roots."
 236 a 220 c 159 g 177 t

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 3,04e-06 Length: 792
 Score: 141.50 Matches: 65
 Percent Similarity: 36.16% Conservative: 33
 Best Local Similarity: 23.99% Mismatches: 114
 Query Match: 4.46% Indels: 59
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0506223 (1-792)

QY 360 ThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPheSer 379
 Db 745 AGCGCGGGATTGAATACATGTCGCCATTC-----AGGCCAAAGTGGAA 701
 QY 380 GlyGlyIleThrTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPhe 399
 Db 700 GCGACAGCGGGAATCATATTTCAGCGTGTGGTCTCGGATCACAAAGCGGAATCCTATT 641
 QY 400 LeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThr 419
 Db 640 ATAGGGGACTACTACAGTACTCCACTACT-----GCAAGTGGCAACACATCATGATGAT 587
 QY 420 MetLeuThrAlaSerThrTrpLysArgGlnLeuLeuAsnAsnValLeuHisProGluAsnGly 439
 Db 586 ATGTTTACTTGCCAAACTT-----GAGACTGTC 560
 QY 440 HisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeu 459
 Db 559 TATAGTGTCTGGTGACCCCTGTTCTTCAGTGTGTTTTCACATGGATCAGGACCT 500
 QY 460 -----IleArgThrSerAlaArgAlaGlyTyr 468
 Db 499 CCTGTGGTCAGAGTCTCTAGTATTTTCAACAGATCAATGCTGCTGTAGAAAAGGATG 440
 QY 469 PhePheThrProGluAsnLysLysLeuGlyThrPheIleAlaArgGlyGlnAlaGlyTyr 488
 Db 439 GTATTAGTCTCTATGCGCT-----CTTCTCCTAAGCTTCTCTGGTGGTCAT 395
 QY 489 ThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAla 508
 Db 489 -----

Db 394 GTGGTC-----GGTAATTTTCCACCTCATGAACGATTTTGTCTTGGTGAACC 347
 QY 509 SerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerVal 528
 Db 346 AATAGTGTAAAGAGGATATGAA-----GAAGGCACAGTT 314
 QY 529 LeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThr 548
 Db 313 GGCCTCTGCCCGTCTTATGCGGTGGATGTGGAGAAATTTTCCTCTCATGGGGCCA 254
 QY 549 LeuSerGlyAlaValPheHisAspMet----- 557
 Db 253 TTAGAAGGGGTCTGATTTTGTGATATTATGGCACAGACCTTGGATCTGGTCCAAGTGTTCCT 194
 QY 558 GlyAspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGlyLeuGlyVal 577
 Db 193 GGTGATCTGCTGCGCAAGCTAAACCTGGAAGTGATATGGATGTGGTGGTTGGTATT 134
 QY 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys 597
 Db 133 CGCGTAGACTCGCCCTCGGACCTTTAAGATTAGATATGCTTC---AATGACACAGGG 77
 QY 598 LysIleArgTrpHisIleSerLeuGlyThrArg 608
 Db 76 ACCGGAGGTTTCACTTTGGGTTGGCCTAAGG 44
 RESULT 2
 AV557254/c
 AV557254/LOCUS
 DEFINITION AV557254 Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone S0064a09f 3', mRNA sequence.
 ACCESSION AV557254
 VERSION AV557254.1 GI:8728669
 KEYWORDS EST
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 607)
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 173-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source
 1..607
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="S0064a09f"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 192 a 172 c 113 g 130 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.82e-05 Length: 607
 Score: 129.00 Matches: 49
 Percent Similarity: 40.12% Conservative: 16
 Best Local Similarity: 30.25% Mismatches: 55
 Query Match: 4.07% Indels: 42
 DB: 10 Gaps: 8
 US-09-857-669-2 (1-609) x AV557254 (1-607)

JOURNAL COMMENT
Unpublished (2000)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1 503
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone_lib="cLEM6L23"
/clone_lib="tomato developing/immature green fruit"
/tissue_type="fruit"
/dev_stage="immature green (5-35 days post-anthesis)"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmGadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
harvested at 7 day intervals through 35 dpa. Equal masses
of tissue from each stage were combined (including seeds
and locules) prior to mRNA isolation."
BASE COUNT 105 a 99 c 146 g 153 t
ORIGIN

Alignment Scores:
Pred. No.: 0.000486 Length: 503
Score: 121.00 Matches: 43
Percent Similarity: 38.96% Conservative: 17
Best Local Similarity: 27.92% Mismatches: 64
Query Match: 3.81% Indels: 30
DB: 10 Gaps: 5

US-09-857-669-2 (1-609) x BE459389 (1-503)

Qy 460 IleArgThrSerAlaArgAlaGlyTyrPheThrProGluAsnLysLysLeuGlyThr 479
Db 96 GTCAATCTCGTCTAGAAAGGATGGPATTAGTCTATGCCT----- 140
Qy 480 PheIleIleArgGlyGlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSer 499
Db 141 CTTCTCTAAGCTCTCTGGTGCATGGTCT-----GGTAATTTCCGCCT 188
Qy 500 GlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIle 519
Db 189 CATGAACGATTGTGCTGGTGGACCAATAGTGTAAAGAGGATATGAA----- 236
Qy 520 GlyLeuAlaGlyProAnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeu 539
Db 237 -----GAAGGCACAGTTGCTCTGCGCGTCTTATGCAGTTGGCTGTGGA 281
Qy 540 GluTyrGlnLeuProPheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 282 GAAATTTCTTCCCTCTGATGGGCCACCTAGAAAGGGGCTGATTGTGATATGGCACA 341
Qy 558 -----GlyAspAlaAlaAlaAsnPheLysArgMetLys 568
Db 342 GACCTTGATCTGTCACAGTTCTCTGCTGATCTCTGCGCAAGGCTAAACCTGGA 401
Qy 569 LeuLysHisGlySerGlyLeuGlyValArgTrpPheSerProLeuAlaProPheSerPhe 588
Db 402 ACTGCATATGATGTGGGTGTATTCGCGTAAAGTCGCCACTCGGACCTCTTAAGATTA 461
Qy 589 AspIleAlaTyrGlyHisSerAspLysLysIleArgTrpHis 602
Db 462 GAGTATGCCTTC---AATGACCAAGCGGACCGGAGGTTTCAC 500
RESULT 5
AW191135/c 494 bp mRNA linear EST 30-MAR-2000
LOCUS AW191135 687022D06.x1 687 - Early embryo from Delaware Zea mays cDNA, mRNA
DEFINITION sequence.
ACCESSION AW191135

AW191135.1 GI:6465676

VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
Zea mays.
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 494)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687022 row: D column: 06.
Location/Qualifiers
source
1..494
/organism="Zea mays"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/clone_lib="687 - Early embryo from Delaware"
/tissue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/note="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI
Site_2: EcoRI; Library was prepared by Statagene using
the Uni-ZAP XR system (Stratagene BN937328-12). Clones
were picked by a Q-bot after blue/white selection
(ampicillin resistance) use 100 micrograms/microliter).
Developed from a pool of equal amounts of RNA from
developing embryos sampled at 14, 21, 28 and 35 days after
pollination of the Illinois High Oil Maize Strain Cycle
90. This closed strain has been selected for high oil
concentration for 90 generations and originates from the
1890s era open pollinated variety Burr's White"

BASE COUNT 147 a 147 c 97 g 103 t
ORIGIN

Alignment Scores:

Pred. No.: 0.00083 Length: 494
Score: 119.00 Matches: 45
Percent Similarity: 38.46% Conservative: 15
Best Local Similarity: 28.85% Mismatches: 64
Query Match: 3.75% Indels: 32
DB: 10 Gaps: 6

US-09-857-669-2 (1-609) x AW191135 (1-494)

Qy 465 ArgAlaGlyTyrPheThrProGluAsnLysLysLeuGlyThrPheIleIleArgGly 484
Db 493 AGCGAGGGCTATGAAATTTGGTCTCT-----GCTAGGCTTCTTCTAAGTGCC 449
Qy 485 GlnAlaGlyTyrThrValAlaArgAspAsnAlaAspValProSerGlyLeuMetPheArg 504
Db 448 TCTGAGCTCACATG-----GAGGGAATTTCCACCTCATGAGCATTTGCA 401
Qy 505 SerGlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyPro 524
Db 400 ATTGGTGGGACAAATAGCTAAGAGGATATGAA----- 368
Qy 525 AsnGlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuPro 544
Db 367 GAAGGTGCTGTGGCTCTGGCTCTTATGCTGTAGGTAGTGGTGAAGTCTCATGCCG 308
Qy 545 PheThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
Db 307 ATGTTTGGGCCCCGGAAGCGTGGTCTTTTGGAGACATATGCTAGTATGTTGTTTGGT 248

QY 572 -----GlySerGlyLeuGlyValArg-----Trp 579

Db 2149 ATCTTGGCCAAACCCTGACTCAGCTGCGGTTCTGGGGATGCGTAAGAGGGCCCTGGTC 2208

QY 580 PheSerProLeuAlaProPheSerPheAspTleAlaTyrGlyHisSerAspLysLysIle 599

Db 2209 TTTGACCGAGTAACTGAGCTGAAGCAGCAGACAGACTTTTGAACACCCGAATCCCAAGAA 2268

QY 500 ArgTrpHisIleSerLeu 605

Db 2269 CAGTGGTGGCTGAAGCTG 2286

RESULT 7

AY109520

LOCUS AY109520 1829 bp mRNA linear HTC 25-MAY-2002

DEFINITION Zea mays CL30_1 mRNA sequence.

ACCESSION AY109520

VERSION AY109520.1 GI:21213271

KEYWORDS HTC.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 1829)

AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanarey,M., Morgante,M. and Tingey,S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes

JOURNAL Unpublished (2002)

REFERENCE 2 (bases 1 to 1829)

AUTHORS Coe,E.C.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

FEATURES

source

1..1829

/organism="Zea mays"

/db_xref="MaizeDB:631282"

/db_xref="taxon:4577"

/clone="CL30_1"

/clone_lib="Maize Mapping Project/DuPont Consensus Library"

/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 369 a 393 c 442 g 383 t 242 others

ORIGIN

Alignment Scores:

Pred. No.: 0.0109 Length: 1829

Score: 117.00 Matches: 93

Percent Similarity: 34.83% Conservative: 47

Best Local Similarity: 23.13% Mismatches: 168

Query Match: 3.69% Indels: 94

DB: 11 Gaps: 21

US-09-857-669-2 (1-609) x AY109520 (1-1829)

QY 160 GluAsnTrpGlnProVal-----GlySerAspPheAspGluAspSer 174

Db 413 GAAGGGTGGCGAGTCCGTCGACAGGTTTCATGCGCGAGTGGAGTTCCTCCCAACACTT 472

QY 175 TrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrProLeuAla 194

Db 473 CACTTTAATGCCATCGCGCGCCAGGATGGCAGATGGTCAACATGTCGTCGCCCATCGTG 532

QY 195 ---LysLeuGlyAsnThr---ArgAlaAlaValAsnProAspThrAlaThrValAspLeu 212

Db 533 CTCTCCGTGAGGACGCACAGCGAAGGGCCATCCAGGCCGCGGCCACG---CGCGTC 589

QY 213 AsnValValValAspSerGlyArgProIleAla----PheGlyAspPheGluIleThrGly 231

Db 590 GCCGTCTCTGACGACGCGACGCCGCCATCGCGCTCTCAGCGACATTGAGATC----- 643

QY 232 ThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArg-----PheGln 247

Db 644 -----TATAAGNNNACGACATNNNNNACAACCTGCA 694

QY 248 ProGlyThrProTyr----- 252

Db 695 CCTGGATTACCTTATGTTGAGGAGCAATTACCAATGCTGCTGACTGCTTATTGTTGGTGGG 754

QY 253 AspLeuAspLeuLeu-----LeuAspPheGlnGlnAlaLeuGluGlnAsnGlyHisTyr 270

Db 755 GACTTGGAGGTATAGAACCAATCAAGTACATGATGCTCTAGATCAGTAT---CGCGTG 811

QY 271 SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgVal----- 287

Db 812 TCTCCAGCACAGCTGCTGAAGAGTTTGCAGCGCAATGCTGATGTCAGTATTTCGCTTT 871

QY 288 -----ProVal-----LysValSerValThrGluValLysArgHis 299

Db 872 CAGCTTCGCCAATCCAGTACACATGGCATGCTCTCTTATGATCCGACACAGCAACGCT 931

QY 300 LysLeuGluThrGlyIleArgLeuAsp-----SerGluTyrGlyLeuGlyGlyLysIle 317

Db 932 CTCCTTGAGATGGGTATNNNNCCCTGTTCTCTGCTCCATCCACATGGGAGGA----- 985

QY 318 AlaTyrAspTyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTyrAspMet 337

Db 986 -----TTCACAAAAGCAGATGATGTCCTCTTAGTTGGAGAAATG 1024

QY 338 AspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsn 357

Db 1025 AAGCAACATGAGAAAGGTCTTGTAGGAAGGTGCTCTCAACCCAGAAATCACTGTTGTTGCG 1084

QY 358 TyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAla 377

Db 1085 ATCTTCCCTCTCCATGCAATTATGCTGGGCCCACTAGGTGCAG----- 1129

QY 378 PheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAla 397

Db 1130 -----TGGCATGCTAAGGCTCGT-----ATTAATGCT-----GGTGCA 1162

QY 398 GluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuGlyAsnSerHis 417

Db 1163 AATTCTCTATTCTTGGAGGGAT-----CCTGCTGCTATGAGCCAT 1204

QY 418 AlaThr-----MetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnVal 433

Db 1205 CCCACGGAGAAAGGCGACCTCTATGATGCTGACCGGGAAGAGGTTTTCAGGATGGCT 1264

QY 434 LeuHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPhe 453

Db 1265 CTGGCCTCGAGAGGCTCAACATCCTCTCTTCAAGGTG----- 1303

QY 454 LeuSerSerThrAlaLeuIleArgThrSerAlaArgAlaGlyTyrPhePheThrProGlu 473

Db 1304 -----GCTGCATATGACACAAAGCAAAAGAAATGATTTCTTTCATCCATCA 1351

QY 474 AsnLysLysLeuGlyThrPheIleLeuArgGlyGlnAlaGlyTyrThrValAlaArgAsp 493

Db 1352 AGGAAACATGATTTCTCTTC-----ATCTCTGGCACAAAGATCGGCACCTCTTGGCCAGAAC 1408

QY 494 AsnAlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerSerValArgGly 513

Db 1409 CGAGAGAGTNNNNNAGATGTTTATGTGCGCGGTGGCTGGAAGGTACTCGTTGAATAC 1468

QY 514 TyrGlu 515

Db 1469 TATGAC 1474

```

RESULT 8
AW053211/c
LOCUS      588 bp      mRNA      linear      EST 20-SEP-1999
DEFINITION 614073B08.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
mRNA sequence.
ACCESSION  AW053211
VERSION     AW053211
KEYWORDS   EST.
SOURCE     Zea mays.
ORGANISM   Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 588)
AUTHORS   Walbot,V.
TITLE     Zea ESTs from various cDNA libraries sequenced at Stanford
JOURNAL   University
COMMENT   Contact: Walbot V
          Department of Biological Sciences
          Stanford University
          855 California Ave, Palo Alto, CA 94304, USA
          Tel: 650 723 2227
          Fax: 650 725 8221
          Email: walbot@stanford.edu
          Plate: 614073 row: B column: 08.
FEATURES   Location/Qualifiers
            source          1..588
                        /organism="Zea mays"
                        /cultivar="W23"
                        /db_xref="taxon:4577"
                        /clone_lib="614 - root cDNA library from Walbot Lab"
                        /tissue_type="root"
                        /dev_stage="3-4 days old"
                        /lab_host="XLOLR"
                        /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
                        EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                        Lab (LM)"
BASE COUNT  182 a 156 c 112 g 138 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0103      Length:      588
Score:          111.00      Matches:      36
Percent Similarity: 42.22%      Conservative: 21
Best Local Similarity: 26.67%      Mismatches: 54
Query Match:     3.50%      Indels:      24
DB:              10      Gaps:      5

US-09-857-669-2 (1-609) x AW053211 (1-588)
QY  476 LysLeuGly---ThrPheLeuLeuArgGlyGlnAlaGlyThrValAlaArgAspAsn 494
|||||  |||||||  |||||||  |||  |||  |||  |||  |||  |||  |||
Db  573 AAGTTGGGTCGCACATTTTGGTTAGTATTGTCACAGCGGATCCATTTGGGGGACATG 514
QY  495 AlaAspValProSerGlyLeuMetPheArgSerGlyGlyAlaSerValArgGlyTyr 514
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  513 GCACCTATCAAGCT-----TTTGCATAGGTGGACTGTAGTGTCCGAGGCTAT 463
QY  515 GluLeuAspSerLleGlyLeuAlaGlyProAsnGlySerValLeuProGluArgAlaLeu 534
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  462 -----GGTAGGGCGGCTGTGGTTCAGGAAGACTATGT 430
QY  535 LeuValGlySerLeuGluTyrGlnLeuProPheThrAlaThrLeuSerGlyAlaValPhe 554
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  429 CTAATTGCTAACTACGAATATACGATTCCTTGGTGAAGCACTCGAGGGTCTATTTC 370
QY  555 HisAspMetGlyAspAlaAlaAlaAsnPheLysArgMet-----LysLeuLys 570
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  369 ATGACACTGGGGAGTACTAGGCTGTGCTGCATGTCCTGGTGAATCCAGCTCTCGCG 310
QY  571 HisGlySer-----GlyLeuGlyValArgTrpPheSerProLeu 583

```

```

||||
Db  309 CAGGGAAACACGAGATTGGTATTGGATTGGATACGACTTCACCTTCAACACTGACCTA 250
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY  584 AlaProPheSerPheAspIleAlaTyrGlyHisSerAspLysLys 598
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  249 GGTGAGATCCGTGTCGACTACGCTATGAATGATGATTTAAGCCGAAA 205
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 9
AW108685/c
LOCUS      503 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays PCO127797 mRNA sequence.
ACCESSION  AW108685
VERSION     AW108685.1 GI:21211868
KEYWORDS   HTC.
SOURCE     Zea mays.
ORGANISM   Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 503)
AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
          Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
          Overgo Probes
JOURNAL   Unpublished (2002)
REFERENCE  2 (bases 1 to 503)
AUTHORS   Coe,E.C.
TITLE     Direct Submission
JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
          Missouri, Columbia, MO 65211, USA
FEATURES   Location/Qualifiers
            source          1..503
                        /organism="Zea mays"
                        /db_xref="MaizeDB:637445"
                        /db_xref="taxon:4577"
                        /clone_lib="Maize Mapping Project/DuPont Consensus
                        Library"
                        /note="this sequence is part of a project of EST
                        assemblies resulting from the application of public
                        contigs to seed DuPont contigs; this resource was
                        assembled by DuPont as part of a collaboration for the
                        overgo addressing of BACS in conjunction with the Maize
                        Mapping Project"
BASE COUNT  151 a 137 c 99 g 116 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0124      Length:      503
Score:          109.50      Matches:      38
Percent Similarity: 42.24%      Conservative: 11
Best Local Similarity: 32.76%      Mismatches: 43
Query Match:     3.45%      Indels:      24
DB:              11      Gaps:      4

US-09-857-669-2 (1-609) x AW108685 (1-503)
QY  506 GlyGlyAlaSerSerValArgGlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsn 525
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  496 GTCGGCAATAAGTAGTGAAGAGGATATGAA-----GAA 464
QY  526 GlySerValLeuProGluArgAlaLeuLeuValGlySerLeuGluTyrGlnLeuProPhe 545
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  463 GGTGCTGTGGCTCTGGCGCTGTATGCTAGGTAGTGTGCTGAAGTGTCTATGCCGCGCATG 404
QY  546 ThrArgThrLeuSerGlyAlaValPheHisAspMet----- 557
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  403 TTTGGCCACTCGAAGCGGTGGTCTTTGGAGACTATGCTAGTGTCTGGTTCGGTCTCT 344
QY  558 -----GlyAspAlaAlaAlaAsnPheLysArgMetLysLeuLysHisGlySerGly 574
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  343 AAGATTCTGTGTGACCCAGCGGAGCTCGTGGAAAGCCAGGAGTGGCTTTGGCTATGCT 284

```

Qy	575	LeuGlyValArg-TrrPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHI	594	PUBMED
Db	283	GTTCGGTCCCGGTGACTCCCTCGGACCTACTACGGCTTGAG-----TACGCCCTT	230	REFERENCE
Qy	594	sSerAspLysIleSylle---ArgTrpHisIleSerLeuGlyThrArg	608	AUTHORS
Db	229	TAATGATAGGCAAGCAAGCAGATTTCACCTTTGGTGTGGCTACAGA	184	
RESULT	10			
AK004789				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

134	QY	IleAlaAsnValGlyValAlaIleLeuGlyAspIle---LeuSerAspGlyAsnLeuAla15	15	QY	LeuSerAspGlyAsnLeuAla15
52	DB	CTCGCGGGCAGGCTCCGACGCGACGCGACATCCGGTGCCTCAAGGGCGTGA	111	DB	CTCGCGGGCAGGCTCCGACGCGACGCGACATCCGGTGCCTCAAGGGCGTGA
153	QY	GluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGln-ProVal-----166	166	QY	GluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGln-ProVal-----166
112	DB	GAATACCGAGAAACTTCCTG-----TGGAAAAGTCTCTATTTGTGCAGATCTTATAAT	165	DB	GAATACCGAGAAACTTCCTG-----TGGAAAAGTCTCTATTTGTGCAGATCTTATAAT
167	QY	-----GlySerAspPheAspGlnAspSerTrpGluAsnse178	178	QY	-----GlySerAspPheAspGlnAspSerTrpGluAsnse178
166	DB	CCCTCAGTGGGACAAAGTACTCATGGCAGGACCTAGATCGGATCGATTGGGGATCAGC	225	DB	CCCTCAGTGGGACAAAGTACTCATGGCAGGACCTAGATCGGATCGATTGGGGATCAGC
178	QY	rlYsthrSerValLeuGlyAlaValThrArglysGlyTyrProLeuAlaIysLeu-----196	196	QY	rlYsthrSerValLeuGlyAlaValThrArglysGlyTyrProLeuAlaIysLeu-----196
226	DB	AANGAACCCAGGTTT-----TTTCAAAGAAGAGTTCCTCATCATGACCCCTCAGATT	279	DB	AANGAACCCAGGTTT-----TTTCAAAGAAGAGTTCCTCATCATGACCCCTCAGATT
197	QY	-GlyAsnThrArgAlaLaValAsnProAspThrAlaThrValAspLeuAsnValVala216	216	QY	-GlyAsnThrArgAlaLaValAsnProAspThrAlaThrValAspLeuAsnValVala216
280	DB	TCAAATACCTGGAGTGAACCGAACCGTCAGAAAGAGGATACGCTTGTC-----CA333	333	DB	TCAAATACCTGGAGTGAACCGAACCGTCAGAAAGAGGATACGCTTGTC-----CA333
216	QY	lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPr236	236	QY	lAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrPr236
334	DB	CCAGAACCCAGGCTTTG-----GACCCCAAGCCACAGAGGCTGAC378	378	DB	CCAGAACCCAGGCTTTG-----GACCCCAAGCCACAGAGGCTGAC378
236	QY	oGluGlnIleValSerGlyLeuAlaArgPheGln-----ProGlyThrProTy252	252	QY	oGluGlnIleValSerGlyLeuAlaArgPheGln-----ProGlyThrProTy252
379	DB	CAAGGAGAGATGCCAATCAAGAAGCAGTTCTCTCACTAGAGGCTCCAGGTTCCCAAG438	438	DB	CAAGGAGAGATGCCAATCAAGAAGCAGTTCTCTCACTAGAGGCTCCAGGTTCCCAAG438
252	QY	rAspLeuAspLeuLeu-----AspPheGlnGlnAlaLeuGlu--265	265	QY	rAspLeuAspLeuLeu-----AspPheGlnGlnAlaLeuGlu--265
439	DB	AGAACTCGCTCTCATCTCTCGGACTCGAGGCTGAAGGGTTTCAGAC-AC	497	DB	AGAACTCGCTCTCATCTCTCGGACTCGAGGCTGAAGGGTTTCAGAC-AC
266	QY	-----GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnI284	284	QY	-----GlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnI284
498	DB	GCACACGGGTGTCGAGAGACCCATGCCGCCAGTTAGCCGGATGTGGAGCTGAGACCTTC557	557	DB	GCACACGGGTGTCGAGAGACCCATGCCGCCAGTTAGCCGGATGTGGAGCTGAGACCTTC557
284	QY	yAspArgValProValIysValSerVal--ThrGluValIysArgHisIysLeuGluTh303	303	QY	yAspArgValProValIysValSerVal--ThrGluValIysArgHisIysLeuGluTh303
558	DB	CAGCAAAACCTCTCTCCAGAGCATAGATCCAGGTTGGATAGGCAT-----606	606	DB	CAGCAAAACCTCTCTCCAGAGCATAGATCCAGGTTGGATAGGCAT-----606
303	QY	rGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIysIleAlaTyrAspTyrTyra323	323	QY	rGlyIleArgLeuAspSerGluTyrGlyLeuGlyGlyIysIleAlaTyrAspTyrTyra323
607	DB	-----CTTCGTAAAGAACTGGATTGGCC-----GTTGTCCCAAGTAATATGC650	650	DB	-----CTTCGTAAAGAACTGGATTGGCC-----GTTGTCCCAAGTAATATGC650
323	QY	nLeuPheAsnIysGlyTyrIleGlySerValValTrpAspMetAspIysTyrGluThrTh343	343	QY	nLeuPheAsnIysGlyTyrIleGlySerValValTrpAspMetAspIysTyrGluThrTh343
651	DB	CTTGAAATCTCGAATACCAAGGCCAGTTGTTGTTGGAAGACTCTTAAGAAAGCGCTCC710	710	DB	CTTGAAATCTCGAATACCAAGGCCAGTTGTTGTTGGAAGACTCTTAAGAAAGCGCTCC710
343	QY	rLeuAlaLaGlyIleSerGlnProArgAsn-----TyrArgGI356	356	QY	rLeuAlaLaGlyIleSerGlnProArgAsn-----TyrArgGI356
711	DB	AGTGTTCGCTCAATCAGGTTTTTCGCTAATAAAGCCAAATATTATCCACAGTTTCCAAAG770	770	DB	AGTGTTCGCTCAATCAGGTTTTTCGCTAATAAAGCCAAATATTATCCACAGTTTCCAAAG770

JOURNAL COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10804 row: i column: 19
High quality sequence stop: 605.
Location/Qualifiers
1. 928
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone IMAGE:4888026
/clone_lib="NICHD XCC Emb2"
/tissue_type="embryo, stage 17/19"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCMV-SPOrt6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb. Constructed by Life Technologies."

FEATURES

source

Alignment Scores:
Pred. No.: 0.172 Length: 928
Score: 103.50 Matches: 72
Percent Similarity: 36.48% Conservative: 40
Best Local Similarity: 23.45% Mismatches: 89
Query Match: 3.26% Indels: 106
Db: 14 Gaps: 17

US-09-857-669-2 (1-609) x B0723872 (1-928)

Qy 98 AsnValysThrMetLeuArgSerLysGlyTyrPheSerLysValSerLeuThrGlu 117
Db 34 AACATCTGTCGCGTCATGAGGCGCAGG ----- 63
Qy 118 LysAspGlyAlaTyrThrValHisIleThrProGlyProArgThrLysIleAlaAsnVal 137
Db 64 -----ATCGAGCCAGGCGCAGAGACTTACGTGGCTCTTTA 99
Qy 138 ---GlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAlaGluTyrTyrArg 156
Db 100 ACCGCATTTGCCGAGAAGGGCATATT-----AACATATCAAAGAG----- 141
Qy 157 AsnAlaLeuGluAsnTrpGlnGlnProValGlySerAspPheAspGlnAspSerTrpGlu 176
Db 142 ---ACCTTAGAAACCTGGAAGAAGTAAGCAAGCCTCACAGCCAGAGATTGATGCAA 198
Qy 177 AsnSerLysThrSerValLeuGlyAlaValThrArgLysGlyTyrPro----- 192
Db 199 -----GTCATCTGGAGTCTGGCTAAGGCTGCATACCCCTCAGTATGTCAG 243
Qy 193 -----LeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThrAlaThrVal 210
Db 244 GACATCATGGCGCATGAGATATGACCGAGATACATA---CCAGATGCGATGAACCTC 300
Qy 211 AspLeuAsnValValAspSerGlyArgProIleAlaPhe----- 224
Db 301 TGTTTGAGTCTGTGACACAAGTTATGAAGATGATGTTTCTTCTAGTGTGTTGAAATCCTC 360
Qy 225 -----GlyAspPheGluIle 229
Db 361 TCGGCTCGCTCGTGCAGATTCGAGAACGGCGATTCTCTACAGCAGCGGAACCTTCTCTG 420
Qy 230 ThrGlyThrGlnArgTyrProGlnGlnIleValSerGlyLeuAlaArgPheGlnProGly 249
Db 421 CGTCACTGCGTCAGTTTG---GAGCAGCCAGCAAGTAAGGTCAAGCATCTCTGTGTGGT 477

Qy 250 -----ThrProTyrAspLeu----- 254
Db 478 TTGAAGAAAGCCAATTTGCATTCGGCACCTCTGCAGTTTCTGTCTGTCTGTCTGTCTGTGGAT 537
Qy 255 -----AspLeuLeuLeuAspPheGlnGlnAlaLeuGluGlnAsnGly----- 268
Db 538 GCCAAAAAACAGATTTAGCTAGATCTCATGAAGTTGCTGAAGAAGAGAGATGTCCT 597
Qy 269 -----HisTyr-SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAs 285
Db 598 GTCAGGCCACACTACTGCTGGCCTCTGCTGATCAGCTTCC---AAAAAGAAAAAATTC 654
Qy 285 pArgValProValLysValSerValThrGluValLysArgHisLysLeuGluThrGlyI 305
Db 655 GAAGGTACCCATTAAAGTAATT-----AAAGCGCTGCTCGAATGGAGT 699
Qy 305 eArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyrTyrAsnLeuPh 325
Db 700 GGAGCTAGACGTTGAG-----ACCTACTCCCACTATGGCCCTG--- 736
Qy 325 eAsnLysGlyTyrIleGlySerValTrpAspMetAspLysTyrGluThrThrLeuAl 345
Db 737 -----AGTGTGTTAATGATGTACAGACGGGCTCGTGCCACA---GC 774
Qy 345 aAlaGlyIleSerGlnPro 351
Db 775 TGCAGGAGAAAGGTCGCC 793

RESULT 15

AI363592/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 423
/organism="Onchocerca volvulus"
/db_xref="taxon:6282"
/clone="SMOVL2CAS01E04"
/clone_lib="Onchocerca volvulus L2 larvae cDNA (SAW98MLW-OV1.2)"
/dev_stage="L2"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 9,000 L2s isolated from infected mosquitoes from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.3 x 10E4 independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr. S.A. Williams, email: genomsmith.edu." 99 a 119 c 111 g 90 t 4 others

BASE COUNT

Wed Nov 13 10:43:47 2002

[illegible]

SOURCE Vitis vinifera.
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 451)
AUTHORS Cramer, G.R. and Cushman, J.C.
TITLE An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay
JOURNAL Unpublished (2002)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer
Plate: 030 row: A column: 02
Seq primer: T3 20mer
High quality sequence stop: 451.
FEATURES
source
1. .451
/organism="Vitis vinifera"
/db_xref="taxon:29760"
/clone_lib="VWA030A02"
/clone_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"
/dev_stage="juvenile and adult"
/tissue_type="leaf"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site.1: EcoRI; Site.2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."
BASE COUNT 92 a 89 c 120 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0998 Length: 451
Score: 101.50 Matches: 35
Percent Similarity: 41.44% Conservative: 11
Best Local Similarity: 31.53% Mismatches: 36
Query Match: 3.20% Indels: 29
DB: 13 Gaps: 5
US-09-857-669-2 (1-609) x BM438066 (1-451)
Qy 513 GlyTyrGluLeuAspSerIleGlyLeuAlaGlyProAsnGlySerValLeuProGluArg 532
|||||
Db 1 GGTATGAA-----GAAGTGCTGTGGGCTCAGGTCGC 33
Qy 533 AlaLeuLeuValGlySerLeuGluTyrGlnLeuProPheThrArgThrLeuSerGlyAla 552
::: |||||
Db 34 TCTCATGTGGTGGTAGTACGAGAAATTTCTTCCCTTGTACGGGGCCATTAGGAGGGCT 93
Qy 553 ValPheHisAspMet-----GlyAspAlaAla 561
::: |||||
Db 94 CTTTTTCTGACATGACACCGATCTTGATCAGGCCCACTGTGCTGGGATCCTGCA 153
Qy 562 AlaAsnPhelysArgMetLysLeuLysHisGlySer-----GlyLeuGlyVal 577
::: |||||
Db 154 GCG-----GCCAGGCTAAAGCCCTGGAAGTGGATATGCGATGTTGGAATT 201
Qy 578 ArgTrpPheSerProLeuAlaProPheSerPheAspIleAlaTyrGlyHisSerAspLys 597
||| |||||
Db 202 CGACTGGACTCTCCTTAGGGCTCTGCCACTTGAATGCTAATGATGCTAATGACCGACGCA 261
Qy 598 LysIleArgTrpHisIleSerLeuGlyThrArg 608
::: |||||
Db 262 CAG---AGGTTTCACTTGGGTTGGTCCACCA 291

RESULT 18
BQ876247
LOCUS BQ876247
DEFINITION AGENCOURT_8742755 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6391121 5', mRNA sequence.
ACCESSION BQ876247
VERSION BQ876247.1 GI:22268253
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1031)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLNL3878 row: n column: 18
High quality sequence start: 103
High quality sequence stop: 510.
FEATURES
source
1. .1031
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="IMAGE:6391121"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site.1: EcoRV; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 249 a 240 c 294 g 246 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.356 Length: 1031
Score: 101.50 Matches: 93
Percent Similarity: 32.76% Conservative: 40
Best Local Similarity: 22.91% Mismatches: 129
Query Match: 3.20% Indels: 145
DB: 14 Gaps: 22
US-09-857-669-2 (1-609) x BQ876247 (1-1031)
Qy 125 HisIleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAsp 144
||| ::|||
Db 31 CACACCTCTCCT---CCAAAAAAGCTGGAGCGGCG-----CGGAAT 75
Qy 145 IleLeuSerAspGlyAsnLeuAlaGluTyrTrpArgAsnAlaLeuGluAsnTrpGlnGln 164
||| ::|||
Db 76 ATTCGAACGAGCATACCCCTCAGGAATTTTAC-----CACTACTATCAGCTT 123
Qy 165 proVal-----GlySerAspPheAsp 171
|||
Db 124 CCAGTCTGTCTCTGAGAGATCCGCCACAAAGCTTAGTTGGGTGAAGTTCTGAT 183
Qy 172 GlnAsp-SerTrpGluAsnSerLysThr---SerValLeuGlyAlaValThrArgLysG1 190
||| |||||
Db 184 CGGACCCCAATGCGCCGAAATCTTTGTACGAGATCCGCTTTCGGAGAGATGTCGAGAGAGA 243
Qy 190 yTyrProLeu-----AlaLysLeuGlyAsnThrArgAlaLalaValAsnProAspTh 207
::: :::
Db 244 ATTCTGTGCATGTCAGCTCAGTCTCTGCACAGCTGGAGCAGCTGCCGCCCATTTGAG 303

```

Qy 207 rAlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPh 227
Db 304 GAATTGATTACTTGGTGT---GTCGTCAT---GATTT 338
Qy 227 eGluIleThrGlyThrGlnArgTyrProGluGlnIleValSerGlyLeuAlaArgPheG1 247
Db 339 GCCAATCCGTGGTGTGTGGCTACATGAAGAG-----AGTGGC-----TTTCCT 383
Qy 247 nProGlyThrProTyrAspLeuAspLeuLeuAspPheGlnClnAlaLeuGluGlnAs 267
Db 384 GCCA-----CA 389
Qy 267 nGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAsp---ArgLeuGlnGlyAspAr 286
Db 390 TAGCCACAAGATAGGCTCTGCACACATTTTGGACTTCCACCTAGAAATTCATGGAGATCG 449
Qy 286 gValPro---ValLysValSerValThrGluValLysArgHisTysLeuGluThrGlyL1 305
Db 450 AATTATATTGGCAATGTTTCGTCGGGAGCTCAAGCCCAACAGCTTGGAT---GGGCT 506
Qy 305 eArgLeuAspSerGluTyrGlyLeu----- 313
Db 507 GCGGCTGTGAATTTCTGGGCTTTACTACACATTTTACGTGTGGCTGGCTCTGAACACG 566
Qy 314 -----GlyGlyLysIleAlaTyrAspTyrTyrAsnLeuPheAsnLysG1 328
Db 567 TCCAGTGGGAGATCCGGAAGTGCAGACGCCGCCCAAGGTGGAT-----GG 614
Qy 328 yTrIleGlySerValValTyrAspMetAspLysTyrGluThrThrLeuAlaAlaGlyL1 348
Db 615 ATGGGAGGC-----CTTCTTTTCCCT 638
Qy 348 eSerGlnProArgAsnTyrArgGlyAsnTyrTrp-----ThrSerAsnValSerTyrAs 366
Db 639 CCAAAACCCCTGGGAAAAATCCCTTTGGGGTGGGTGGGTCGCCCAATTTCCCATTT 698
Qy 366 nArgSerThrThrGlnAsnLeuGluLysArgAlaPheSerGlyGlyIleTrpTyrValAr 386
Db 699 CAAAACCTTTTCCCATGTTTGGGGGGGGGCGCTTTTGTGGGGGGGTTTTTTTATTAA 758
Qy 386 g-----AspArgAlaGlyIleAspAlaArgLeuGlyAlaGluPheLeuAlaGluGlyAr 404
Db 759 CACTTGGGGAGGGGGGGGTTTTTTTGGGGGGGGG----- 798
Qy 404 gLysIleProGlySerAspIleAspLeuGlyAsnSerHisAlaThrMetLeuThrAlaSe 424
Db 799 -----CCCGGGGCTAAATTTCCCAAAAATTTG----- 828
Qy 424 rTrpLysArgGlnLeuLeuAsnValLeuHisProGluAsnGlyHisTyrLeuAspG1 444
Db 829 -----CCCGGGGGGTCCCTTTTCCCAA 854
Qy 444 yLysIleGlyThrThrLeuGlyThrPheLeuSerSerThrAlaLeuIleArgThrSerAl 464
Db 855 AAAATATGGGACCCCTTTTGGGACTTTTCCCGGGGAACC----- 894
Qy 464 aArgAlaGlyTyrPhePheThrPro-----GluAsnLysLysLeuGlyTh 479
Db 895 -----ACACCCCTCTTAAACCCGGAATAAAAAAACCAGCCGAGC 935
Qy 479 rPheIleIleArgGly 484
Db 936 GTTTCCTCTGGGGGG 951

```

RESULT 19
 BH770975/c 3095 bp DNA linear GSS 01-MAY-2002
 LOCUS LLMCTag700 MGI363 Random Sequence Tag Library Lactococcus lactis
 DEFINITION subsp. cremoris genomic, DNA sequence.
 ACCESSTION BH770975
 VERSION BH770975.1 GI:20373932
 KEYWORDS GSS.
 SOURCE Lactococcus lactis subsp. cremoris.

```

ORGANISM Lactococcus lactis subsp. cremoris
REFERENCE 1 (bases 1 to 3095)
AUTHORS Bolotin.A., Ehrlich.S.D. and Sorokin.A.
TITLE Studies of genomes of dairy bacteria Lactococcus lactis
JOURNAL Sci. Aliments. (2002) In press
COMMENT Contact: Sorokin A
Genetique Microbienne
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologues in strain IL1403 is rpoC (92%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 3067.
Location/Qualifiers
source
1..3095
/organism="Lactococcus lactis subsp. cremoris"
/db_xref="taxon:1359"
/clone_lib="MGI363 Random Sequence Tag Library"
/note="Vector: pSGM2; Site_1: SmaI; Library of
chromosomal fragments of L.lactis strain MGI363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 915 a 649 c 614 g 912 t 5 others
ORIGIN
Alignment Scores:
Pred. No.: 1-93 Length: 3095
Score: 101.50 Matches: 104
Percent Similarity: 34.55% Conservative: 67
Best Local Similarity: 21.01% Mismatches: 187
Query Match: 3.20% Indels: 137
DB: 17 Gaps: 26
US-09-857-669-2 (1-609) x BH770975 (1-3095)
Qy 33 GlyPheAlaLeuPheLysAsnLysSerProAspThrGluSerValLysLeuLysPro--- 51
Db 2583 GGGATTCAAGCTTTTGAACCAAGTCTTAATGTGTAAGCCATCCGCTTTCACCCACTT 2524
Qy 52 -----LysPheProVal 55
Db 2523 GCTGTGAAGCCTATAACGCCGACTTTTGACGGTGACCAATGGCCATTCACCTCCCAT 2464
Qy 56 ArgIleAspThrGln---AspSerGluLleLysAspMetValGluGluHisLeu----- 72
Db 2463 TCTGAAGAAGCACCACTGAAGCAGCTCTTTTGATGCTGTGCTGAACATATCTTGAAC 2404
Qy 73 -----ProLeuIleThrGlnGlnGlnGluValLeuAspLys----- 85
Db 2403 CCTAAAGATGGTAACCACTGTTACACCATCTCAAGATATGGTCTTGGTAACTACTAC 2344
Qy 86 -----GluGlnThrGlyPheLeuAlaGlu-----GluAlaProAspAsn 98
Db 2343 CTTACAAATGGAAAGAAAGCGTGAAGCGGAAGGAATGATTTTGGCACTCTCTGAAGAA 2284
Qy 99 ValLysThrMetLeuArgSerLysGlyTyrPheSerSerLysValSerLeuThrGluLys 118
Db 2283 GTTGAATCGCAATGCGTAAC---GCTTAT-----GTACACTTCATACACGT 2239
Qy 119 AspGlyAlaTyrThrValHisIle-----ThrProGlyProAspThrLysIle 134
Db 2238 ATTTGGTATCCGCAAAATCACTCAATAAACCTTCGCTGAATAATCAACAAGATAAATC 2179
Qy 135 -----AlaAsnValGlyValAlaIleLeuGlyAspIleLeuSerAspGlyAsnLeuAla 152
Db 2178 TTGGTAACCACTCTTGGTAAAGTTATTTTCAACTCAATCATCTCCAGAAGGA-----ATG 2125
Qy 153 GluTyrTyrArgAsnAlaLeuGlu---AsnTrpGlnGlnProValGlySerAspPheAsp 171

```

```
Db 2124 CCTACTTGAACGAACCAACTGATGTTAACTTGACAACCTTCAACTGATGACCGCTCTTT 2065
|||
|||
Qy 172 GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThr----- 187
|||
|||
Db 2064 ATGGATGCTGTGCTCAAAACATCAAGAAAGTTTTCGCTGGAACTCATCACTGTCGCCATTC 2005
|||
|||
Qy 188 ArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThr 207
|||
|||
Db 2004 AAAAAGAGATAT-----CTTGGAAATATTATCGCGGAAGTA----- 1969
|||
|||
Qy 208 AlaThrValAspLeuAsnValValAspSerGlyArgProIleAlaPheGlyAspPhe 227
|||
|||
Db 1968 -----TTCAAAAGGTTTTC 1957
|||
|||
Qy 228 GluIleThrGlyThrGlnArgTyrProGluGlnIle-----ValSer 241
|||
|||
Db 1956 CGTACACAGCTACATCTGAGTACCTTGACCGTTTCAAGAGACTTGGTTACCACCAATCT 1897
|||
|||
Qy 242 GlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln 261
|||
|||
Db 1896 ACTTTGGCTGTTTGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1840
|||
|||
Qy 262 GlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArg 281
|||
|||
Db 1839 GAAATCATTTGATGCT---GCCATAAACCGGTAGACACATCACTAAACCAATTCGCTGCC 1783
|||
|||
Qy 282 LeuGlnGlyAspArgValProValLysValSerValThrGluValLysArgHisLysLeu 301
|||
|||
Db 1782 -----GGTTTGATTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1750
|||
|||
Qy 302 GluThrGlyIleArgLeuAspSerGluTyrGlyLeuGlyLysIleAlaTyrAspTyr 321
|||
|||
Db 1749 GTTACTGCTGTTGGCTGATGCAAAAGACTCTCTCGAAAGAGTATTGTTGAAGC--- 1693
|||
|||
Qy 322 TyrAsnLeuPheAsnLysGlyTyrIleGlySerValValTrpAspMetAspLysTyrGlu 341
|||
|||
Db 1692 ---CAAGAATTGTCACAA-----CCCATCGTTATGATGATGATGATGATGATGATGATGAT 1648
|||
|||
Qy 342 ThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyrArgGlyAsnTyrTrpThrSer 361
|||
|||
Db 1647 CGTGCTAATATTTCACACTCTCTCAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1588
|||
|||
Qy 362 Asn-----ValSerTyrAsnArgSerThrThrGlnAsnLeu 373
|||
|||
Db 1587 AATGTAATAATCATGGAATGCTTATCATCTCAACTTCCGTGAAGGCGTTTCTGCTCTTG 1528
|||
|||
Qy 374 Glu-----LysArg 376
|||
|||
Db 1527 GAAATGTTCTTCTCAACTCACGGTCTGCTGAAGGGATGCCGATACGGCCCATAGACA 1468
|||
|||
Qy 377 AlaPheSerGlyGlyIleTyrTyrValArgAspArgAlaGlyIle-----AspAla 393
|||
|||
Db 1467 GCCGACTCAGGT-----TACCTTACCTGCTGTTGGTTGCTGCTGCCAAGATGTT 1417
|||
|||
Qy 394 ArgLeuGlyAlaGluPheLeuAlaGluArgLysIleProGlySerAspIleAspLeu 413
|||
|||
Db 1416 ATCATTCGTGAAAGAGGATGTTGGTACTGACCGTGAACTTGATGCTGATATTGCAACT 1357
|||
|||
Qy 414 GlyAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeu----- 429
|||
|||
Db 1356 GTAAAGAA-----ATGGTTGAACCACTCTTTGAACCGTTGGTGGCTGTACACT 1306
|||
|||
Qy 430 LeuAsnAsnValLeuHisProGluAsnGlyHisTyrLeuAspGly 444
|||
|||
Db 1305 CGTAAATCAGTCTCTATCCAGAAACTGGTGAATGATTATTGGC 1261
|||
|||
RESULT 20
BQ872838
LOCUS BQ872838 724 bp mRNA linear EST 15-AUG-2002
DEFINITION OG116M16.yg.ab1 OG-ABCDI lettuce salinas Lactuca sativa cDNA clone
BQ116M16, mRNA sequence.
ACCESSION BQ872838
```

VERSION
KEYWORDS
SOURCE
ORGANISM

BQ872838.1 GI:22259397

EST
Lactuca sativa

REFERENCE
AUTHORS

Kozik A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
Lin, H., Van Damme, N., Lavelle, B., Chevalier, P., Ziegler, J., Ellison,
P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, F., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compendomics.ucdavis.edu/
Unpublished (2002)

JOURNAL
COMMENT

Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmunsen Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-742-9659
Email: akozik@catc.org [michelmore@vegmall.ucdavis.edu]
belongs to contig OG_CA_Contig6253, see http://cgdb.ucdavis.edu/
for details.
Plate: OG116 row: M column: 16.

FEATURES
source

Location/Qualifiers
1..724
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="OG116M16"
/clone_lib="OG-ABCDI lettuce salinas"
/lab_host="E.coli"
/note="vector: pBRCNDSF1AB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG_116-OG-ABCDI lettuce salinas
TAG_TISSUE-flowers pre-fertilized
TAG_SEQ-GCCTGACGGG"
BASE COUNT 255 a 126 c 147 g 196 t
ORIGIN

Alignment Scores:

Pred. No.: 0.315 Length: 724
Score: 100.00 Matches: 46
Percent Similarity: 40.18% Conservative: 42
Best Local Similarity: 21.00% Mismatches: 87
Query Match: 3.15% Indels: 44
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ872838 (1-724)

Qy 26

LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45

Db 3

GTAACGAGGCGCAACGACAGCATCATCTCCATCCATCAAG-----GGTACTGAG 53

Qy 46

ServallysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65

Db 54

TCAGCCAGC-----CAAGCTTTAGACACAAACAATCT----- 86

Qy 66

AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLys 85

Db 87

-----GTGTGGTAGTATTCATCATCTTGGTGATACAGCTCATCTCATCACCANA 140

Qy 86

GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105

Db 86

-----GTGTGGTAGTATTCATCATCTTGGTGATACAGCTCATCTCATCACCANA 140

Db 141 CAAACTGAC-----ATGGCAAAATCAATTGATTGATACACGAGACTTCTTCTGTCAGGAG 194

QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125

Db 195 AAGGTATATCGGTGCTATCATTGATGACCAAGTGTGGATCAATCATCCAA 254

QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145

Db 255 ATGGGTCTGATTGACTAGCAGCAAGCATTAGAAACAACCAAGTCGGTGTGCGAGGTATC 314

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165

Db 315 CCAATTGTTGGATGCGAGAG-----TCTATTATTACACACAAACAGAC 362

QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184

Db 363 ATGGCTAGTGTAGTCGGTGTGAAGCTCAACTTCGTCGAGCAAGCATGGTTATTAGGA 422

QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203

Db 423 GCTATTATTACAGCGGACGAGCAGGTGTAGAAGATGTTGGAACACA----- 470

QY 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219

Db 471 -----GCAACAAAATCAATTGGAAGATATCCCAATTCTAGGATCT 506

QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232

Db 507 GAAAAACCTAGAGAAACACATCCAGCATTTGGAGATATCCCAATTCTAGGATCT 563

RESULT 21

BQ871231

LOCUS

DEFINITION

QGI11D13.yg.abl OG.ABCDI lettuce salinas Lactuca sativa cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Alexander Kozik [R.W.Michelmor]

Department of Vegetable Crops, R.W.Michelmor Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmor@vegmail.ucdavis.edu]

belongs to contig QG_CA_Contig6253, see <http://cgdb.ucdavis.edu/> for details.

Plate: QGI11 row: D column: 13.

FEATURES

source

1..613

Location/Qualifiers

organism="Lactuca sativa"

cultivar="Salinas"

db_xref="taxon:4236"

clone="QGI11D13"

clone_lib="OG.ABCDI lettuce salinas"

lab_host="E.coli"

note="vector: pBRCDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgdb.ucdavis.edu/>

TAG_LIB-OG.ABCDI lettuce salinas

TAG_TISSUE-flowers pre-fertilized

TAG_SEQ-GCTTGACGG*

BASE COUNT 206 a 116 c 129 g 162 t

ORIGIN

Alignment Scores:

Pred. No.: 0.324 Length: 613

Score: 99.00 Matches: 46

Percent Similarity: 40.18% Conservative: 42

Best Local Similarity: 21.00% Mismatches: 87

Query Match: 3.12% Indels: 44

DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ871231 (1-613)

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45

Db 3 GTAAAGGAGGACACACAGGATACATCCATCCAAAG-----GGTACTGAG 53

QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65

Db 54 TCAGTCAG-----CAAGCTTTAGACACAAACAATCT----- 86

QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluValLeuAspLys 85

Db 87 -----GTTCGTTGCTAGTATTCCAATCATTCGATACACTCAATCCATCATCCAA 140

QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105

Db 141 CAAACTGAC-----ATGGCAAAATCAATTGATGATACACGAGCTTCTTCTGTCAGGAG 194

QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125

Db 195 AATGGTATATTCGGTCTATCATTGACTCAGATGTTGGTGGATCAATCATCCAA 254

QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145

Db 255 ATGGGTCTGATTGACTAGCAGCAAGCATTAGAAACAACCAAGTCGGTGTGCGAGGTATC 314

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165

Db 315 CCAATTGTTGGAGACGACAGAG-----TCTATTATTACACACAAACAGAC 362

QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184

Db 363 ATGGCTAGTGTAGTCGGTGTGAAGCTCAACTTCGTCGAGCAAGCATGGTTATTAGGA 422

QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203

Db 423 GCTATTATTACAGCGGACGAGCAGGTGTAGAAGATGTTGGAACACA----- 470

QY 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219

Db 471 -----GCAACAAAATCAATTGGAAGATGTTGGAAGATGTTGATCTGTC 506

QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232

Db 507 GAAAAACCTAGGAGAAACACATCCAGCATTTGGAGATATCCCAATTCTAGGATCT 563

RESULT 22

BQ871231

LOCUS

DEFINITION

QGI11D13.yg.abl OG.ABCDI lettuce salinas Lactuca sativa cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Alexander Kozik [R.W.Michelmor]

Department of Vegetable Crops, R.W.Michelmor Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmor@vegmail.ucdavis.edu]

belongs to contig QG_CA_Contig6253, see <http://cgdb.ucdavis.edu/> for details.

Plate: QGI11 row: D column: 13.

FEATURES

source

1..613

Location/Qualifiers

organism="Lactuca sativa"

cultivar="Salinas"

db_xref="taxon:4236"

clone="QGI11D13"

clone_lib="OG.ABCDI lettuce salinas"

lab_host="E.coli"

note="vector: pBRCDNASFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that

BM560625 1171 bp mRNA linear EST 20-FEB-2002

AGENCOURT_6597970 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:543221

5', mRNA sequence.

BM560625

BM560625.1 GI:18805199

from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>. TAG_LTB-OC ABCDI lettuce salinas TAG_TISSUP-flowers pre-fertilized TAG_SEQ-GCTTCACGGG*

BASE COUNT 209 a 116 c 130 g 167 t
ORIGIN

Alignment Scores:

Pred. No.: 0.439 Length: 622
Score: 98.00 Matches: 47
Percent Similarity: 40.47% Conservative: 40
Best Local Similarity: 21.86% Mismatches: 92
Query Match: 3.09% Indels: 36
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0875918 (1-622)

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAACGAGGACACAGCAGGTACATCATCCAAAG-----CGTACTGAG 53
QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86
QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLys 85
Db 87 -----GTTGTGGTAGTNTCCAACTATGTCATACAGCTCAATCCATCATCACAAA 140
QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAAACTGAC-----ATGGCAATATGATTGATACCCAGCACTTCTCTCAAGGAG 194
QY 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
Db 195 AATGCTATATTCGGTCTCACTTCACTCACTCAGATGTTGGTGGATCATCATCAAA 254
QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATCAGCTAGCAGCAGCATTAGAAACACCAAGTCGGTTCGCGAGGTATC 314
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnPro 165
Db 315 CCAATTGTTGGATGCGCAGAG-----TCATTATTACACACAAACAGAC 362
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATGCGGTTGAAAGCTCAACTTCGCTCGGTCAAGAAGCATGGTTTATTAGA 422
QY 185 AlaVal-----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
Db 423 GCTATTTTATCAGCGCGCAGCAGAGTTGTAGAGATGTTGGAACAC----- 470
QY 204 AsnProAspThrAlaThrValAspLeuValValValAspSerGlyArgProIleAla 223
Db 471 -----GCAACAAATAATCATTTGAAAGAGTTCTGATTCTCTGTC 506
QY 224 PheGlyAspPheGluIleThrGlyThrGln-----ArgTyrPro 236
Db 507 GAAAAACCCCTAGAGAAAAACCACTCCAGCATTTGGAGATATCCC 551

RESULT 24

BF535617

LOCUS

DEFINITION 602054012F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4193306 5', mRNA sequence.

ACCESSION

BF535617 GI:11622985

VERSION

EST.

KEYWORDS

Mus musculus

SOURCE

house mouse

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 994)

NIH-MSC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM9525 row: g column: 03

High quality sequence stop: 656.

Location/Qualifiers

1..994

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4193306"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:1;

Not1; Site:2: SalI; Cloned unidirectionally; Primer: Oligo

dt. Average insert size 1.3 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 296 a 226 c 296 g 176 t

ORIGIN

Alignment Scores:

Pred. No.: 0.902 Length: 994
Score: 98.00 Matches: 76
Percent Similarity: 34.47% Conservative: 35
Best Local Similarity: 23.60% Mismatches: 124
Query Match: 3.09% Indels: 88
DB: 12 Gaps: 14

US-09-857-669-2 (1-609) x BF535617 (1-994)

QY 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125

Db 5 AGGGGTCGAGGAGCGGAAGCGCGGAGCTTAGTCCGGGAGGTGGAGCGCTGAGCGG 64

QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145

Db 65 TTAACGCTCTACACAGCTTCTCCTCGCTCCGACCTGTCCGCCGCCGCTCTGAGACACC 124

QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGln 164

Db 125 TCGGCTTCTCCCTTCCTTCCCAAGAGTAAACATGACTGAGTGTGCTCATATCT 184

QY 165 ProValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerValLeuGly 184

Db 185 GCTCTGGGAGAAAACCTGTCAAGCAACATGGAGAAA-----CTACAT 229

QY 185 AlaValThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsn 204

Db 230 GCAGCGACCCACCAAG-----AACATAATCTTGCCTCTCT 265

QY 205 -----ProAspThrAlaThrValAspLeuAsnValValAspSerGly 219

Db 266 TCCAGTTCAACATTTCTCGACCTAAAGGTTGGACGTTGGATGCTTGGTT----- 316

QY 220 ArgProIleAlaPheGlyAspPheGluIleThrGlnArgTyrProGluGlnIle 239


```

Db 1546 TCAGCACTGGGACCCACG 1563
    :::: ||||| |||
RESULT 26
BQ871709 630 bp mRNA linear EST 15-AUG-2002
LOCUS QG112J17.yg.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION QG112J17, mRNA sequence.
ACCESSION BQ871709
VERSION BQ871709.1 GI:22258259
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa.
REFERENCE 1 (bases 1 to 630)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
P., Kolkman,J., Stabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL http://compgenomics.ucdavis.edu/
COMMENT Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmunsdon Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA.Contig6253, see http://cgpdb.ucdavis.edu/
for details
Plate: QG112 row: J column: 17.
FEATURES
    source
        Location/Qualifiers
            1..630
                /organism="Lactuca sativa"
                /cultivar="Salinas"
                /db_xref="taxon:4236"
                /clone="QG112J17"
                /clone_lib="QG-ABCDI lettuce salinas"
                /lab_host="E.coli"
                /note="vector: pRCNMF1AB; The library was constructed
                from 10 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_Lib-QG-ABCDI lettuce salinas
                TAG_TISSUE=flowers pre-fertilized
                TAG_SEQ=GGTTGACGGG"
BASE COUNT 213 a 118 c 130 g 169 t
ORIGIN
Alignment Scores:
Pred. No.: 0.593 Length: 630
Score: 97.00 Matches: 45
Percent Similarity: 39.73% Conservative: 42
Best Local Similarity: 20.55% Mismatches: 88
Query Match: 3.06% Indels: 44
DB: 14 Gaps: 9
US-09-857-669-2 (1-609) x BQ871709 (1-630)
Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAAAGGAGCAACAGCAAGGTACATCCATCATCCAAAG-----GGTACTGAG 53
Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65

```

```

Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAAATCT----- 86
Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLys 85
Db 87 -----GTTGTTGGTAGTATTCCAATCATTCAGCTCAATCCATCATCACAAA 140
Qy 86 GluGlnThrGlyPheIleAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAACATGAC-----ATGGCAATCAATTGATTGATACCAGCAGCTTCTTCGCAAGGAG 194
Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
Db 195 AATGGTATATATTCGGTGCATCACTTCAATTCAGTCAAGATGCTGTCGATCAATCAACAA 254
Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATTCACTAGCAGCAAGCATTAGAACACCAACGCGTTGTCGGAGGTATC 314
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 315 CCAATTTGTTGGGAATGCAGCAGAA-----TCTATTATTACACACAAACAGAC 362
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATCGCGTTGAAAGCTCAACTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 422
Qy 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaIleVal 203
Db 423 GCTATTTTATCAGCGCGCAGCAGCGTTGTCAGAAGATGTTGGAACACA----- 470
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
Db 471 -----CCAAACAAATCATTTGAAAAGGTTCTGATTCGTCTGTC 506
Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 507 GAAAAAACCTAGAGAAAACACACATCCAGCATTCGAGATATCCCAATTCAGGATCT 563
RESULT 27
BQ479447 590 bp mRNA linear EST 03-JUN-2002
LOCUS Ku36h01.y1 Strongyloides ratti PA female naive S11 TOFO v1
DEFINITION Strongyloides ratti cDNA similar to SW1MYL1_CAEEL P98080
HYPOTHETICAL 51.7 KD PROTEIN F56D2.1 IN CHROMOSOME III. [1] ; mRNA
sequence.
ACCESSION BQ479447
VERSION BQ479447.1 GI:21315414
KEYWORDS EST.
SOURCE Strongyloides ratti.
ORGANISM Strongyloides ratti.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Pangrolaimoidea; Strongyloidea; Strongyloides.
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Claire Murphy and Dr. James McCarter
at Washington University, St. Louis. Parasitic adult females were
collected from naive animals and provided by Dr. Mark Viney of

```


US-09-857-669-2 (1-609) x B0872562 (1-577)

```

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAAACGAGCACAACAGGTCATCTCCATCATCCAAAG-----GGTACTGAG 53
Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLys 65
Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAATAATCT----- 86
Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
Db 87 -----GTTGTTGGTAGTATCCATCATCTGGTATACAGCTCAATCATCATCCAA 140
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAAACTGAC-----ATGGCAAACTAATTGATTCACACAGCACTTCTCTGTCAGAGGAG 194
Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
Db 195 AATGGTATATTCGGTCTATCACTTCTATGACTCAGAAATGTTGGTGGATCAATCATCCAA 254
Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATTTCAGCTAGCCGAGCAGCATTAGAAACAACCAAGTCGGTTGTCGAGGTATC 314
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 315 CCAATTGTTGGGAATCCAGCAGAA-----TCTATTATTACACACAAACAGAC 362
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATCGCGTTCAAGCTCAACTCGTGGTCAAGAGCATGTTTATTAGGA 422
Qy 185 AlaVal----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 423 GCTATTATTCAGCGCGAGCAGCAGGTTGTAGAAGATGTGGAAACACA----- 470
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
Db 471 -----GCAAAACAATAATCATTTGAAAAGGTTCTGATTCTGTCT 506
Qy 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 507 GAAAAAACCTAGAGAAACCACTCCAGCATTCGAGATATCCCAATTTCTAGGATCT 563

```

B0875018 660 bp mRNA linear EST 15-AUG-2002
 LOCUS OG16020.yg.abl.0G.ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION OG16020.mRNA sequence.
 ACCESSION B0875018
 VERSION B0875018.1 GI:22261578
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Lactuceae; Lactuceae;
 Lactuca.

REFERENCE 1 (bases 1 to 660)
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 TITLE Lactuca and Sunflower ESTs from the Compositae Genome Project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmudson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA.Contig96253, see http://cgpdb.ucdavis.edu/
 for details.
 Plate: QG16 row: 0 column: 20.

FEATURES

Location/Qualifiers
 1..660
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone_lib="OG-ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="Vector: pBRCDNASFIAB: The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_LfIB-OG-ABCDI lettuce salinas
 TAG_TISSUE=flowers pre-fertilized
 TAG_SEQ=GCCTTGACGGG"

BASE COUNT 222 a 121 c 136 g 180 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.844 Length: 660
 Score: 96.00 Matches: 46
 Percent Similarity: 40.57% Conservative: 40
 Best Local Similarity: 21.70% Mismatches: 96
 Query Match: 3.03% Indels: 30
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0875018 (1-660)

```

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
Db 3 GTAAACGAGCACAACAGGTCATCTCCATCATCCAAAG-----GGTACTGAG 53
Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLys 65
Db 54 TCAGCCAGC-----CAAGCTTTAGACACACAATAATCT----- 86
Qy 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
Db 87 -----GTTGTTGGTAGTATCCATCATCTGGTATACAGCTCAATCATCATCCAA 140
Qy 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
Db 141 CAAACTGAC-----ATGGCAAACTAATTGATTCACACAGCACTTCTCTGTCAGAGGAG 194
Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
Db 195 AATGGTATATTCGGTCTATCACTTCTATGACTCAGAAATGTTGGTGGATCAATCCAA 254
Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
Db 255 ATGGGTTCTGATTTCAGCTAGCCAGCAGCATTAGAAACAACCAAGTCGGTTGTCGAGGTATC 314
Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
Db 315 CCAATTGTTGGGAATCCAGCAGAA-----TCTATTATTACACACAAACAGAC 362
Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
Db 363 ATGGCTAGTATCGCGTTCAAGCTCAACTCGTGGTCAAGAGCATGTTTATTAGGA 422
Qy 185 AlaVal----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 423 GCTATTATTCAGCGCGAGCAGCAGGTTGTAGAAGATGTGGAAACACAATAATCT 482

```

```

QY 204 -----AsnProAspThrAlaThrValAspLeuAsnValValAspSerGlyArg 220
DB 483 ATTGAAAGGTTCTGATCTGCGGAAAAACCTTAGAGAAACCACATCCAGC----- 536
QY 221 ProIleAlaPheGlyAspPheGluIleThrGlyThr 232
DB 537 -----ATTGCAGATATCCCAATCTAGGATCT 563

RESULT 30
BQ875030
LOCUS QG16P11.yg.ab1 QG-ABCD1 lettuce salinas Lactuca sativa cDNA clone
DEFINITION QG16P11.yg.ab1 QG-ABCD1 lettuce salinas Lactuca sativa cDNA clone
ACCESSION BQ875030
VERSION BQ875030.1 GI:22261590
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
1 (bases 1 to 677)
Kozik A., Micheltore R.W., Knapp S., Matvienko M., Rieseberg L.,
Lin H., van Damme M., Lavelle D., Chevalier P., Ziegler J., Ellison
P., Kolkman J., Slabaugh M.S., Livingston K., Zhou Y., Lai Z.,
Church S., Jackson L. and Bradford K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Micheltore]
Department of Vegetable Crops, R.W.Micheltore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [micheltore@vegmail.ucdavis.edu]
Belongs to contig QG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
for details.
Plate: QG16 row: P column: 11.
FEATURES
source
1..677
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone_lib="QG-ABCD1"
/lab_host="E.coli"
/notes="Vector: pBRCDNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_LIB-QG-ABCD1 lettuce salinas
TAG_TISSUE=flowers pre-fertilized
TAG_SEQ=CGTTGAGCGG"

BASE COUNT 231 a 122 c 140 g 184 t
ORIGIN

Alignment Scores:
Pred. No.: 0.878 Length: 677
Score: 96.00 Matches: 45
Percent Similarity: 39.73% Conservative: 42
Best Local Similarity: 20.55% Mismatches: 88
Query Match: 3.05% Indels: 44
DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x BQ875030 (1-677)

```

```

QY 26 LeuSerGluAsnLysAlaIleGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
DB 3 GTAAACGAGGACCAAGCAGCATCATCATCCATCCAAAG-----GGTACTCAG 53
QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIleLys 65
DB 54 TCAGCCAGC-----CAAGCTTTAGACACAACAATCT----- 86
QY 56 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGluGluValLeuAspLys 85
DB 87 -----GTTGTTGGTAGTATCCCAATTCAGTATACAGTCTGATACATCCACCAA 140
QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
DB 141 CAATCTGAC-----ATGGCAATCAATTTGGTTGATACCACGACTTCTCTGTCAAGCAG 194
QY 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
DB 195 AATGGTATATTCGGTGTCTATCATTCATTCAGTCAGAAATGTTGGTGGATCAATCCAG 254
QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
DB 255 ATGGGTTCTGATTCAGCTAGCAAGCATTTAGAAACAACCAAGTCGTTGTCGAGGTATC 314
QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
DB 315 CCAATTGTTGGGAATGACGACAGAA-----TCTATTATTACACACACACACAGAC 362
QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
DB 363 ATGGCTAGTATGCGGTTGAAGCTCAACTTCGTCGGTCAACAGCATGGTTATTAGGA 422
QY 185 Alaval---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaVal 203
DB 423 GCTATTATTACGACGACGACGAGGTGTAGAGATGTTGGAAACACA----- 470
QY 204 AsnProAspThrAlaThrValAspLeuAsnValValAspSerGly----- 219
DB 471 -----GCAACAAATCATTTGAAAGAGTTCTGATTCTGTC 506
QY 220 -----ArgProIleAlaPheGlyAspPheGluIleThrGlyThr 232
DB 507 GAAACAAACCTTAGAGAAACACATCCAGCATTTGGAGATATCCCAATTTCTAGGATCT 563

RESULT 31
BQ874536
LOCUS QG15M18.yg.ab1 QG-ABCD1 lettuce salinas Lactuca sativa cDNA clone
DEFINITION QG15M18.yg.ab1 QG-ABCD1 lettuce salinas Lactuca sativa cDNA clone
ACCESSION BQ874536
VERSION BQ874536.1 GI:22261177
KEYWORDS EST.
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
Lactuca.
1 (bases 1 to 731)
Kozik A., Micheltore R.W., Knapp S., Matvienko M., Rieseberg L.,
Lin H., van Damme M., Lavelle D., Chevalier P., Ziegler J., Ellison
P., Kolkman J., Slabaugh M.S., Livingston K., Zhou Y., Lai Z.,
Church S., Jackson L. and Bradford K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Micheltore]
Department of Vegetable Crops, R.W.Micheltore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659

```

Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig6253, see http://cgpub.ucdavis.edu/
 for details.
 Plate: QG15 row: M column: 18.

FEATURES
 source
 Location/Qualifiers
 1..731
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QG15M18"
 /clone_lib="QG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpub.ucdavis.edu/
 TAG_LIB-QG_ABCDI lettuce salinas
 TAG_TISSUE-flowers pre-fertilized
 TAG_SEQ-GCTTGACGGG"
 BASE COUNT 260 a 128 c 146 g 197 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.988 Length: 731
 Score: 96.00 Matches: 47
 Percent Similarity: 40.36% Conservative: 81
 Best Local Similarity: 21.08% Mismatches: 52
 Query Match: 3.03% Indels: 10
 DB: 14 Gaps: 10

US-09-857-669-2 (1-609) x B0874536 (1-731)

QY 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 Db 3 GTAAAGGAGGACAAAGCAAGGATACATCCATCATCCAAAG -----GGTACTCAG 53
 QY 46 SerValLysLeuLysProLysPheProValArgIleAspThrClnAspSerGluLeuLys 65
 Db 54 TCAGCCAG -----CAAGCTTTAGACACACAAATCT----- 86
 QY 66 AspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluValLeuAspLys 85
 Db 87 -----GTTGTTGGTAGTATTTCCCAATCATTTGGTGATACAGCTCAATCCATCATCCACAA 140
 QY 86 GluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArgSer 105
 Db 141 CAACCTGAC -----ATGGCAATCAATGATTGATACACAGCACTTCTTCTGTCAAGGAG 194
 QY 106 LysGlyTyrPheSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AATGGTATATTTCGGTGCTATCATCTTCACTCAGAAATGTTGGTGATCAATCATCCAA 254
 QY 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGGTCTGATTACAGTAGCCAGCAAGCAATAGAACACCAACCAAGTCGGTGTGGAGGTATC 314
 QY 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 315 CCAATTGTTGGGAATGCAGAGAA -----TCATTATTACAACACAA----- 356
 QY 166 ValGlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrSerVal----- 182
 Db 357 -----ACAGACATGGGTAGAGATCGGGTTGAAAGCTCAACTTCGTCGTCGAAGAGCAT 410
 QY 183 -----LeuGlyAlaVal-----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThr 199
 Db 411 GGTATTATAGGAGCTATTTTATCAGCGCAGGACAGGCTGTAGAACATGTTGGAAACACA 470

OY 200 ArgAlaAlaValAsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly 219
 Db 471 -----GCAACAAATCATTCATGAAAAAGT 494
 QY 220 -----ArgProileAlaPheGlyAspPheGluIle 229
 Db 495 TCTGATTCTGTGAAAAAACCCCTAGAAAAACACACATCCAGCATTCGAGATATCCCAATT 554
 QY 230 ThrGlyThr 232
 Db 555 CTAGGATCT 563
 RESULT 32
 B0850925
 LOCUS QG13024.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION QG13024, mRNA sequence.
 ACCESSION B0850925
 VERSION B0850925.1 GI:22236394
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.
 REFERENCE 1 (bases 1 to 735)
 AUTHORS Kozik A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevaller, P., Ziegler, J., Ellison
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.
 Letteuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 JOURNAL Unpublished (2002)
 COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig6253, see http://cgpub.ucdavis.edu/
 for details.
 Plate: QG13 row: O column: 24.
 FEATURES
 Location/Qualifiers
 1..735
 /organism="Lactuca sativa"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QG13024"
 /clone_lib="QG_ABCDI lettuce salinas"
 /lab_host="E.coli"
 /note="vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpub.ucdavis.edu/
 TAG_LIB-QG_ABCDI lettuce salinas
 TAG_TISSUE-flowers pre-fertilized
 TAG_SEQ-GCTTGACGGG"
 BASE COUNT 259 a 128 c 149 g 199 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.996 Length: 735
 Score: 96.00 Matches: 46
 Percent Similarity: 40.57% Conservative: 40
 Best Local Similarity: 21.70% Mismatches: 96
 Query Match: 3.03% Indels: 30

[illegible]

```

Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [micheltmorevegmail.ucdavis.edu]
belongs to contig QG_CA_Contig6253, see http://cgpdb.ucdavis.edu/
for details
Plate: QG112 row: D column: 16.
      Location/Qualifiers
        l..757
        /organism="Lactuca sativa"
        /cultivar="Salinas"
        /db_xref="taxon:4236"
        /clone="QG112D16"
        /clone_lib="QG-ABCO1 lettuce salinas"
        /lab_host="E.coli"
        /note="Vector: pBRCDNASFIAB: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu"

```

BASE COUNT	250 a	134 c	159 g	204 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.04	Length:	757	
Score:	96.00	Matches:	45	
Percent Similarity:	39.73%	Conservative:	42	
Best Local Similarity:	20.55%	Mismatches:	88	
Query Match:	3.03%	Indels:	44	
DB:	14	Gaps:	9	

US-09-857-669-2 (1-609) x B0871570 (1-757)

Ov 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45

DB 33 GT'AACGAGGCACAGGCACAGGTACATCCATTCATCCAAAG-----GGTACTGAG 83

0v 46 serValIysIleuIysproIyspheproValArgIleAspThrGlnAspSerGluIleIys 65

84 TC ACC ACC-----CAGCGCTTTGACACACACACAAATCT-----TAA

[illegible][illegible][illegible][illegible][illegible]

.....

[illegible][illegible][illegible]

DD 343 CCAT TGT TGGGAA TCGACAGCA
 ----- TCTAT TAT TACACACACACACACACAC

```

QY 100 VAGTYSEABPNHABPGRINABPSEIFRPGUASH--$ELYSIINSEFVILEGRY 18
      :: ||||| :: ||| ||| ::|||
      :: ||||| :: ||| ||| ::|||

```

DB 393 ATGGCTAGTGTATGCGGTCTGAAAGCTCAACTTCGTCTGGTCAAGAAGCATGGTTTATTAGGA 452

QY 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 20

Db 453 GCTATTTATCAGCCGCAGGCTGTAGAAGATGTTGGAACACA-----50


```

Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
Db 501 -----GCAAACAATAATTCGAAAGAGTTCTGATTCGTGC 536
Qy 220 -----ArgProTleAlaPheGlyAspPheGluIleThrGlyThr 232
Db 537 GAAAAACCTAGAGAAACACATCCAGCATTCGAGATATCCCAATTCAGGATC 593

RESULT 34
BM919327
LOCUS
DEFINITION
  BM919327 1072 bp mRNA linear EST 12-MAR-2002
  5', mRNA sequence.
ACCESSION
  BM919327
VERSION
  BM919327.1 GI:19369706
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 1072)
REFERENCE
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: Life Technologies, Inc.
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Place: LLNL12776 row: 1 column: 17
  High quality sequence stop: 572.
  Location/Qualifiers
    1..1072
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:5748568"
      /clone_lib="NIH_MGC_120"
      /lab_host="DH10B"
      /note="Organ: pooled pancreas and spleen; Vector:
      pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
      source anonymous pool of spleen and pancreas from 28 yo
      male. Library is oligo-dT primed and directionally cloned
      (EcoRV site is destroyed upon cloning). Average insert
      size 1.5 kb, insert size range 1-2.5 kb. Library is
      normalized and enriched for full-length clones and was
      constructed by C. Gruber (Invitrogen). Research Genetics
      tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT
  270 a 340 c 292 g 170 t
ORIGIN

Alignment Scores:
  Pred. No.: 1.78 Length: 1072
  Score: 96.00 Matches: 79
  Percent Similarity: 33.91% Conservative: 39
  Best Local Similarity: 22.70% Mismatches: 116
  Query Match: 3.03% Indels: 115
  DB: 14 Gaps: 16

US-09-857-669-2 (1-609) x BM919327 (1-1072)

Qy 5 ProThrAlaLeuLeuProAlaLeuPhePheProHISAlaTyAlaProAlaAla 24
Db 6 CCAAGCGCCTTCGCCACCGAGGAG-----CCCCAC-----CCACCAAGGCC 50
Qy 25 AspLeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThr 44
Db 51 AAGATGTCCAGCAAGGGGCCCAACCAAGACCAACCAAGAGCGGCCACACGGGCCACA 110

```

```

Qy 45 GluSerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluIle 64
Db 111 TCCAAATGTC-----TTGCATATGTTTACCAGTCCCGACATCCAGGAGTTT 155
Qy 65 LysAspMetValGluGluHisLeuProLeuIleThrGlnGlnGlnGluGluValLeuAsp 84
Db 156 AAG-----GAGGCTTTCAACATGATTGACCAGAACCGGTGATGCTTCATTGAC 203
Qy 85 LysGluGlnThrGlyPheLeuAlaGluAlaProAspAsnValLysThrMetLeuArg 104
Db 204 AAGGAG-----GACCTGACGACATGCTGGGCC 230
Qy 105 SerLysGly-----TyrPheSerLysValSerLeuThrGluLys 118
Db 231 TCGCTGGGGAAGAACCCACAGCAATACCTGGGAGGCATGATGAC-----GAGGC- 283
Qy 119 AspGlyAlaTyrThrValHisIleThrPro----- 128
Db 284 CCGGGGGCCCATCAACTTCACCATGTTCTCCATGTTTGGGGAAGCTGAACGGCAC 343
Qy 129 GlyProArg-ThrLysIleAlaAsnValGly----- 138
Db 344 GGACCCGAGGATGTGATTGCAACGCTTTGCGCTTCTTCGACGAGGAGCCTCAGGTTT 403
Qy 139 -----ValAlaIleLeuGlyAspIleLeuSerAspG1 149
Db 404 CATCCATGAGGACACCTCCGGAGCTGCTCACCACCATGGGTGACCGCTTCACAGATGA 463
Qy 149 yAsnLeuAlaGluTyrTyArgAsnAlaLeuGluAsnTrpGlnGlnProValGlySerAs 169
Db 464 GGAAGTGGAGAGATGTACCGGGAGGCA-----CCCATTCATAAGAA 505
Qy 169 pPheAsp-GlnAspSerTrpGluAsnSerLysThrSerValLeuGlyAlaValThrArgL 189
Db 506 AGCAACTTCAACTACGTGGAGTTCCCGCGATCTCTCAACATGGCG----- 552
Qy 189 ySGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaValAsnProAspThraLa- 208
Db 553 -----CCAAGGATAAAGACGACTAGGGCACCCCGCCCTTCACACCCAG 598
Qy 209 -----ThrValAspLeuAsnValValV 216
Db 599 CCCCAGCAGTACCCCTCCCGCACACCCGTCATACCATGCTCCCTGCCATGACC 658
Qy 216 aLAspSerGlyArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrP 236
Db 659 TCCTCAGGATCCCTCC-----TTTGGGGGTAGGTGTCACAGTCCAGTCCAG 703
Qy 236 roGluGlnIleValSerGlyLeuAlaArgPhe----- 246
Db 704 TGGAAAAACAGGGCCAGGAGAAAGTGGTGGCCCAAGCTTGAGGCAGATGTTTCCACAG 763
Qy 247 --GlnProGlyThrProTyrAspLeuAspLeuLeuAspPheGln-----G 262
Db 764 GTGACCCCGAGACCCCTGGGGCTATAGTTCTCTCACCCTTCCCAAGGAAAAAGACCCA 823
Qy 262 lAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPheAspArgL 282
Db 824 CCTTCTTGGGGAACATGGGCGCTGGGAAGGGGAGGAAC-----TTTAAAGGCA 874
Qy 282 euGlnGlyAspArgValPro 288
Db 875 CCAAGGGGAAGGGGCCCC 894

RESULT 35
BM975465
LOCUS
DEFINITION
  BM975465 963 bp mRNA linear EST 15-JUL-2002
  5', mRNA sequence.
ACCESSION
  BM975465
VERSION
  BM975465.1 GI:21786299
KEYWORDS
  EST.
SOURCE
  human.

```

```

US-09-857-669-2 (1-609) x B0675465 (1-963)

Qy 14 PhePheProHisAlaTyrAlaProAlaAlaAspLeuSerGluAsnLysAlaAlaGly 33
    ::::: |||
Db 94 TATTTCTTCACATCGTGCCGGCGCGGCCGGAGAGCCGCCTCATCGCGGTGGC 153
    ::::: |||
Qy 34 PheAlaLeuPheLysAsnLysSerProAspThrGlnSerValLysLeuLysProLysPhe 53
    ::::: |||
Db 154 TAGGTG-----CAGCACGCGAGTTC----- 174

Qy 54 ProValArgIleAspThrGlnAspSerGluIleLysAspMetValGluGluHisLeuPro 73
    ::::: |||
Db 175 ---GTGCGGTTCGACAGGAGCGCGCGAGCCAGAAG----ATGGAGCGCGGGCGCG 225
    ::::: |||
Qy 74 LeuIleThrGlnGlnGluGluValLeuAspLysGlnThrGlyPheLeuAlaGlu 93
    ::::: |||
Db 226 TGGATAGACCAGAGGGCGCGGAGTAGTTGGACCGAGGACACGGGAATTGAGAGGCCCA 285

Qy 94 GluAlaProAsp-----AsnValLysThrMetLeuArgSerLysGlyTyrPheSer 111
    ::::: |||
Db 286 TCACAGACTGACCGAGCGAAGCTGGGAGCCGTG-----CGCGGCTACTACAACCA 336
    ::::: |||
Qy 112 LysValSerLeuThrGluLysAspClyValaTyrThrValHisIleThrProGlyProArg 131
    ::::: |||
Db 337 -----AGCGAGGACGGTCTCCACACCATCCAGATAATGATGTGCG----- 375

Qy 132 ThrLysIleAlaAsnValGlyValAlaIleuGlyAspIleLeuSerAspGlyAsnLeu 151
    ::::: |||

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.fhlr.gov
Plate: LLAM10726 Row: C Column: 08

Rate: ERM10/20 IOW: C Column: 08
High quality sequence stop: 753

Location/Qualifiers

10861-10862

```

I. : I086
/organism="Homo sapiens"

```

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
/ab_xref="taxon:9606"
/cloc="IMACE:4830571"
```

```
/clone="IMAGE:4820671"
/clone lib="NIH MCC 95"
/clone
```

```
/clone_lib="NIH_MGC_95"
```

```

/tissue_type="hippocampus"

```

```

/label_host="DH108"
/notes="Organ: brain; Vector: pBluescriptPR (modified
pBluescript KS+); Site.1: BamHI; site.2: SalI-XhoI (atcgaag
); Oligo-dr primed using primer 5'-TTTCTTTTCTTTTCTTTT-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carnanci, in preparation). Library
constructed by M. Brownstein (NIMH/NGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
179 a 341 c 239 t 2 others

```

INSTITUTES OF HEALTH). NOTE: THIS IS A N
179 a 341 c 325 a 239 f 2 others

DATE OF ORIGIN

Alignment Scores:

Argument Scores:		
Pred. No.:	2.09	Length: 1086
Score:	95.50	Matches: 76
Percent Similarity:	34.85%	Conservative: 39
Best Local Similarity:	23.03%	Mismatches: 97
Query Match:	3.01%	Indels: 118
DB:	12	Gaps: 17

US-09-857-669-2 (1-609) x BG704719 (1-1086)

Qy	269	HlSty	SerGlyAlaSerValGlnAlaAspPheAspArgLeuGlnGlyAspArgValPro	288
Db	240	CACCATCCGGACCTGG-	-----GGCAAAAGAAGCCCA	272
Qy	289	ValLysVal-SerValThrGluValLysArgHisLysLeuGlu--	---ThrGlyIleArgLeu	307
Db	273	CTTCTACTACAACATCTCTCAGGTGAAGGTCACAGAGTCGCAACTGACATCTTCCGAGCT	332	
Qy	307	uasp-----	-----SerGluTyrGlyLeuLg	314
Db	333	CGATTTCAGCCACAGCAGGAGTGATGCTTCAAATCAACATGCTCTTGGGGGTGGG	392	
Qy	314	yGlyLysIleAlaTyrAspTyrTrpAsnLeuPheAsnLysGlyTyrIleGlySerValVa	334	
Db	393	CTTCGGGAGACAGCTGCTCTACTGGTCTTCTATGATGGGGCTACATCAACGCC	447	
Qy	334	lTrpAspMetAspLysTyrGluThrThrLeuAlaAlaGlyIleSerGlnProArgAsnTyr	354	
Db	448	-----TCAGCTCAGGGTGTGCCATCGGCACCTGGTCTGGAGCTCTCCCGGATCC	497	
Qy	354	rArgGlyAsnTyrTrpThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuLg	374	
Db	498	CCCTGGCGGATGAAGCTGCCAATGCTCTCCGCGGCCTCTCTCTCCAGA	551	
Qy	374	uLysArgAlaPheSerGlyGlyIleTrpTyrValArgAspArgAlaGlyIleAspAlaAr	394	
Db	552	GCACGCGCCTTCGGGGGAACCTTCAAGAAGGTGATGAT	591	
Qy	394	gLeuGlyAlaGluPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeuLg	414	
Db	592	-----TTTCTC	597	
Qy	414	yAsnSerHisAlaThrMetLeuThrAlaSerTrpLysArgGlnLeuLeuAsnValLe	434	
Db	598	-----TCCAGGTTCAATACC---TCAGGGATGGCGTCTCTCTCAACCAAGCAGAT	644	
Qy	434	uHisProGluAsnGlyHisTyrLeuAspGlyLysIleGlyThrThrLeuGlyThrPheLe	454	

```

/!lab_host="DH10B"
/!note="vector: pONR-LIB"
BASE COUNT      751 a      484 c      500 g      731 t
ORIGIN

Alignment Scores:
Pred. No.:      7.38      Length:      2466
Score:          95.50      Matches:      81
Percent Similarity: 36.44%      Conservative: 56
Best Local Similarity: 21.54%      Mismatches: 125
Query Match:      3.01%      Indels:      114
DB:              11      Gaps:      20

US-09-857-669-2 (1-609) x BC013324 (1-2466)

Qy 74 LeuIleThrGlnGlnGlnGluValLeuAspLys-----GluGlnThrGly 89
Db 418 ATTGCTGAATACCAGCAATCAAGTCTCAGATAGAAATCTGTGTCAGAAATATCTTC 477
Qy 90 PheLeuAlaGluGluAlaProAspAsnValLysThrMetLeuArgSerLysGlyTyrPhe 109
Db 478 ATGCTTGCA-----GGGCTGATTTGAAGGAAGATCTCCAGAAACTTAAAGGTCGCCGA 531
Qy 110 SerSerLysValSerLeuThrGlu-----LysAspGlyAlaTyrThrValHisIle 126
Db 532 ACAGCTGAACCAAAATGACAAAGAGTCAATCTAGCTGCCCGTTCATCATTCAC--- 588
Qy 127 ThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIleLeu 146
Db 589 ACAGCTGGGACCTAAATATATAAAGCGGCTATCCAGCAGAGCT----- 630
Qy 147 SerAspGlyAsnLeuAlaGluAlaGlyTyrArgAsnAlaLeuGluAsnTrpGlnProVal 166
Db 631 ---GAGAGTTCCCTTTATAGCTGCCTACAGAAACGTACTT----- 666
Qy 167 GlySerAspPheAspGlnAspSerTrpGluAsnSerLysThrServalLeuGlyAlaVal 186
Db 667 -----CAACTACAAAGAGCAGTCAATGCTTCTGTGTGCTCTCTGTC 711
Qy 187 -----ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
Db 712 ATCAATTCGCAAAAGCTGGTATCTTTA----- 741
Qy 204 AsnProAspThrAlaThrValAspLeuAsnValVal-ValAspSer-----G1 219
Db 742 ---GAGGATGCAACACACATACACATCCACTCTGAAGAGATTTCCCTAGAGATTCATGGG 798
Qy 219 YArgProIleAlaPheGlyAspPheGluIleThrGlyThrGlnArgTyrProGluGlnI 239
Db 799 GAAACCATTT-----GAAAAAGT 816
Qy 239 eValSerGlyLeuAlaArgPheGlnProGlyThrProTyrAspLeuAspLeuLeuAs 259
Db 817 AGTATTGCTGCTCTGATCTTGAAGAGGGTACT---TACCAAAAGCTGCTACTCTCTA 873
Qy 259 pPheGlnGlnAlaLeuGluGlnAsnGlyHisTyrSerGlyAlaSerValGlnAlaAspPh 279
Db 874 CTTCCCAAGGTCATTAAAGAGAGGAGATCGA---TCATTGCCCTACCTACCTGCAGATAT 930
Qy 279 eAspArgLeuGlnGlyAsp-----ArgValProValLysVa 291
Db 931 TGGAAATGCAAGAGGGGCGCTGTGTTACCTGAACGACAGATATAGATAAGTCAGAAACC 990
Qy 291 lSerValThrGluValLysArgHisLysLeuGluThrGlyIleArgLeuAspSerGluTy 311
Db 991 TGGTGTCTCCAGAGATAAACCAAGAGAGGAGATGAAGGCTTGGAGTTCAT----- 1042
Qy 311 rGlyLeuGlyGlyLysIleAlaTyrAspTyrTrpAsnLeuPheAsnLysGlyTyrIleG1 331
Db 1043 -----CTCTCTTTCATGG 1056
Qy 331 ySerValValTrp-----AspMetAspLysTyrGluThrThrLeuAlaAl 346
Db 1056 -----

```

```

Db 1057 CTCTCATGCTTTTGTCTCGAATGGAAGGAGATATTGACAAAGCAAAAGAACTGATCTTCA 1116
Qy 346 aGly---IleSerGln-----ProArgAsnTyrArgGlyAsnTy 358
Db 1117 GGGACAATTTATCAGAGAGAGCTCTGCAGAAAGCAGCATCAAGAAATTTAT-----AATCG 1170
Qy 358 rTTPThrSerAsnValSerTyrAsnArgSerThrThrGlnAsnLeuGluLysArgAlaPh 378
Db 1171 CTGGTTATGT-----CAAGCAAGATCTGAGGATCTGTCTGATATTGCTTC 1215
Qy 378 eSerGlyGlyIleTTPThrValArgAspArgAlaGlyIleAspAlaArgLeuGlyAlaG1 398
Db 1216 TCTAAAGAGCCTTATAC-----CAACACAGCTTGTGATCACTGTGTCGGAACAGT 1263
Qy 398 uPheLeuAlaGluGlyArgLysIleProGlySerAspIleAspLeu 413
Db 1264 GATCGGTGCTAGTTGGAGAAACATTCCTGTAACATTAATAGATATG 1309

RESULT 38
BQ875852
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN

```

Alignment Scores:

Pred. No.: 0.912 Length: 578
 Score: 95.00 Matches: 45
 Percent Similarity: 39.73% Conservativeness: 42
 Best Local Similarity: 20.55% Mismatches: 88
 Query Match: 2.99% Indels: 44
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0875483 (1-578)

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45
 Db 3 GTAAGAGGACGACAGAGGATCCATCCATCCAAAG-----GGTACTGAG 53
 Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLys 65
 Db 54 TCAGCCAGC-----CAAGCTTTAGACACAAATCT----- 86
 Qy 66 AspMetValGluGluHisLeuProLeuLleThrGlnGlnGluValLeuAspLys 85
 Db 87 -----GTTGTTGGTAGTATCCATCAATCGTCACTCAATCCATCCACCAA 140
 Qy 86 GluGlnThrGlyPheLeuAlaGluCluAlaProAspAsnValLysThrMetLeuArgSer 105
 Db 141 CAAACTGAC-----ATGCAAAATCAATTGATTGATACACGACTTCTGTCAGGAG 194
 Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125
 Db 195 AATGGTATATTCGGTGCATCACTCACTGACTCAGAAATGTTGGTGGATCAATCCAA 254
 Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145
 Db 255 ATGGGTTCTGATTGACTAGCAGCAGCATAGAACACCAAGTCGGTGTGCGAGGTATC 314
 Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165
 Db 315 CCAATTTGTTGGGAATCCAGCAGAA-----TCTATTATTACACACAAACAGAC 362
 Qy 166 ValGlySerAspPheAspGlnAspSerTrpGluAsn---SerLysThrSerValLeuGly 184
 Db 363 ATGGCTAGTATCGCGTTGAAAGCTCAACTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 422
 Qy 185 AlaVal---ThrArgLysGlyTyrProLeuAlaLysLeuGlyAsnThrArgAlaAlaVal 203
 Db 423 GCTATTATTATCATCGCGCAGCAGAGGTTGTAGAGATGTTGGAACACAC----- 470
 Qy 204 AsnProAspThrAlaThrValAspLeuAsnValValValAspSerGly----- 219
 Db 471 -----GCAAAACAAATCAATTCGAAAGGTTCTGATTCTGTC 506
 Qy 220 -----ArgProIleAlaPheGlyAspPheGluLeuThrGlyThr 232
 Db 507 GAAAAACCTAGACAAACCATCCAGCATTCGAGATATCCCAATTCAGGATCT 563

RESULT 39

B0875483

LOCUS

B0875483 581 bp mRNA linear EST 15-AUG-2002

DEFINITION OG18E01.yg.abi OG-ABCDI lettuce salinas Lactuca sativa cDNA clone

ACCESSION B0875483

VERSION B0875483.1 GI:22264704

KEYWORDS EST

SOURCE Lactuca sativa.

ORGANISM

Lactuca sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

1 (bases 1 to 581)

REFERENCE

AUTHORS

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, D., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
 Church, S., Jackson, L. and Bradford, K.

TITLE

JOURNAL

COMMENT

Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]

Belongs to contig OG_CA_Contig6253, see http://cgdb.ucdavis.edu/

for details.

Plate: QG18 row: E column: 01.

FEATURES

source

Location/Qualifiers

1..581

/organism="Lactuca sativa"

/cultivar="Salinas"

/db_xref="taxon:4236"

/clone="OG18E01"

/lab_host="E.coli"

/note="Vector: pBRCDNASFIAB; The library was constructed

from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated,

directionally cloned into a custom medium-copy vector and

transformations made with four size classes to minimize

size bias. Details of each source of RNA and library

construction can be obtained at http://cgdb.ucdavis.edu/

TAG.LIB-OG-ABCDI lettuce salinas

TAG.TISSUE=flowers pre-fertilized

TAG_SEQ=GGTGGAGGG

BASE COUNT 198 a 115 c 122 g 146 t

ORIGIN

Alignment Scores:

Pred. No.: 0.919 Length: 581
 Score: 95.00 Matches: 45
 Percent Similarity: 39.73% Conservativeness: 42
 Best Local Similarity: 20.55% Mismatches: 88
 Query Match: 2.99% Indels: 44
 DB: 14 Gaps: 9

US-09-857-669-2 (1-609) x B0875483 (1-581)

Qy 26 LeuSerGluAsnLysAlaAlaGlyPheAlaLeuPheLysAsnLysSerProAspThrGlu 45

Db 3 GTAAGAGGACGACAGAGGATCCATCCATCCAAAG-----GGTACTGAG 53

Qy 46 SerValLysLeuLysProLysPheProValArgIleAspThrGlnAspSerGluLeuLys 65

Db 54 TCAGCCAGC-----CAAGCTTTAGACACAAATCT----- 86

Qy 66 AspMetValGluGluHisLeuProLeuLleThrGlnGlnGluValLeuAspLys 85

Db 87 -----GTTGTTGGTAGTATCCATCAATCGTCACTCAATCCATCCACCAA 140

Qy 86 GluGlnThrGlyPheLeuAlaGluCluAlaProAspAsnValLysThrMetLeuArgSer 105

Db 141 CAAACTGAC-----ATGCAAAATCAATGATTGATACACGACTTCTTCTGTCAGGAG 194

Qy 106 LysGlyTyrPheSerSerLysValSerLeuThrGluLysAspGlyAlaTyrThrValHis 125

Db 195 AATGGTATATTCGGTGCATCACTTCACTGACTCAGAAATGTTGGTGGATCAATCCAA 254

Qy 126 IleThrProGlyProArgThrLysIleAlaAsnValGlyValAlaIleLeuGlyAspIle 145

Db 255 ATGGGTTCTGATTGACTAGCAGCAGCATAGAACACCAAGTCGGTGTGCGAGGTATC 314

Qy 146 LeuSerAspGlyAsnLeuAlaGluTyrTyrArgAsnAlaLeuGluAsnTrpGlnGlnPro 165

Db 315 CCAATTTGTTGGGAATCCAGCAGAA-----TCTATTATTACACACAAACAGAC 362

Search completed: November 9, 2002, 02:55:17
Job time : 1979 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:20:19 : Search time 14 Seconds
(without alignments)
627.205 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPPHAY.....IAYGHSDKKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues
Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146	4.6	3472	9 US-10-027-806-4	Sequence 4, Appli
2	122	3.8	1781	9 US-09-995-749A-2	Sequence 2, Appli
3	118.5	3.7	815	10 US-09-815-242-5106	Sequence 5106, Ap
4	116	3.7	922	10 US-09-886-468-19	Sequence 19, Appl
5	111	3.5	976	10 US-09-871-212-6	Sequence 6, Appli
6	108.5	3.4	794	10 US-09-864-761-37663	Sequence 37663, A
7	108.5	3.4	1375	10 US-09-740-274-4	Sequence 4, Appli
8	108.5	3.4	3158	10 US-09-815-242-12611	Sequence 12611, A
9	106	3.3	3034	10 US-09-737-149-25	Sequence 25, Appl
10	106	3.3	3034	10 US-09-737-149-25	Sequence 30, Appl
11	104.5	3.3	382	10 US-09-912-020-299	Sequence 299, App
12	102	3.2	2025	10 US-09-815-242-5703	Sequence 5703, Ap
13	101	3.2	982	10 US-09-841-132-176	Sequence 176, App
14	101	3.2	1006	10 US-09-841-132-190	Sequence 190, App
15	100.5	3.2	939	10 US-09-815-242-10797	Sequence 10797, A
16	100	3.2	2086	10 US-09-815-242-5639	Sequence 5639, Ap
17	100	3.2	5795	10 US-09-815-242-12610	Sequence 12610, A
18	99.5	3.1	1025	10 US-09-897-056-7	Sequence 7, Appli
19	99	3.1	764	10 US-09-747-521-4	Sequence 4, Appli

20	99	3.1	764	12	US-10-106-014-4	Sequence 4, Appli
21	99	3.1	915	10	US-09-332-226-2	Sequence 2, Appli
22	98.5	3.1	512	10	US-09-813-820-6	Sequence 6, Appli
23	98	3.1	6281	10	US-09-815-242-12996	Sequence 12996, A
24	97.5	3.1	685	10	US-09-815-242-13477	Sequence 13477, A
25	97.5	3.1	801	10	US-09-815-242-5330	Sequence 5330, Ap
26	97.5	3.1	1055	10	US-09-815-242-12189	Sequence 12189, A
27	97	3.1	514	8	US-08-973-028-4	Sequence 4, Appli
28	97	3.1	711	10	US-09-765-272-82	Sequence 82, Appli
29	97	3.1	1087	10	US-09-918-909-24	Sequence 24, Appli
30	96.5	3.0	778	10	US-09-815-242-10728	Sequence 10728, A
31	96.5	3.0	1215	10	US-09-815-242-5908	Sequence 5908, Ap
32	96.5	3.0	1269	10	US-09-815-242-13113	Sequence 13113, A
33	96.5	3.0	1277	10	US-09-969-362-3	Sequence 3, Appli
34	96.5	3.0	1609	12	US-10-007-693-75	Sequence 75, Appli
35	96.5	3.0	1752	10	US-09-841-132-180	Sequence 180, App
36	96.5	3.0	2437	10	US-09-815-242-5834	Sequence 5834, Ap
37	95.5	3.0	841	10	US-09-815-242-5779	Sequence 5779, Ap
38	95.5	3.0	841	10	US-09-815-242-12751	Sequence 12751, A
39	95.5	3.0	1017	10	US-09-897-056-5	Sequence 5, Appli
40	95	3.0	871	10	US-09-886-468-21	Sequence 21, Appli
41	95	3.0	1430	10	US-09-740-274-6	Sequence 6, Appli
42	94.5	3.0	425	10	US-09-815-242-13263	Sequence 13263, A
43	94.5	3.0	790	10	US-09-801-368-228	Sequence 228, App
44	94	3.0	417	10	US-09-815-242-10492	Sequence 10492, A
45	94	3.0	932	10	US-09-815-242-5578	Sequence 5578, A

ALIGNMENTS

RESULT 1

US-10-027-806-4
; Sequence 4, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOOP 002A
; CURRENT APPLICATION NUMBER: US/10/027, 806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match	4.6%	Score 146;	DB 9;	Length 3472;
Best Local Similarity	21.3%	Pred. No. 0.0021;		
Matches 128;	Conservative 59;	Mismatches 213;	Indels 200;	Gaps 31;
QY	138	GVAIIGDILSDGNLAAYRNALENWQPVGSDFDSDWENSKTSVLGAVTRKGYPLAKLG	197	
Db	2087	GSSGLNAVISHGELDTLSFYLD---IPYGAELDIDRL---LPLVGVPT--GFEEF----	2135	
QY	198	NTRAAVNPDTATVDLVNVVDSGRPIAFGDFEITGTQYRPQIVSGLARFPQGPYDLDL-	256	
Db	2136	-----DNGRQLYIGAFR--DSQSSPGTLPAGLQRYELGIPYDLASA	2174	
QY	257	-----LLDFQ--OALRQNGHYSGASVQAD---FDRLOGDRVPKVSVEYVKRHKLET	303	
Db	2175	VFAQSLGIFDFPPFNGMRANGSLAGLHVPPDGSILFRAGNAERTVISY---DMDSDHLD	2231	
QY	304	-----GIRLSEYGLGKIAY-----DYYNLFNK-----GYTGSVV	334	
Db	2232	LSFRESFKPDVQSQSTPNIR-DMDISPDGMFLYLLQGVLDVNYLTDYSILDAPAYAGTL-	2289	

APPLICANT: Zhang, Linong
APPLICANT: Wu, Qiaohua
TITLE OF INVENTION: MODIFIED BOVINE ADENOVIRUS HAVING
TITLE OF INVENTION: ALTERED TROPISM
FILE REFERENCE: 293102003000
CURRENT APPLICATION NUMBER: US/09/871.212
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 60/208,678
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 976
TYPE: PRT
ORGANISM: Bovine adenovirus 3
US-09-871-212-6

Query Match 3.5%; Score 111; DB 10; Length 976;
Best Local Similarity 22.2%; Pred. No. 0.31; Indels 174; Gaps 36;
Matches 146; Conservative 78; Mismatches 260;

QY 3 IKPTALL----LPALFFPHAYAPADLSEKKAAGFALFKNSPDTSYKLPKF-----53
DB 66 INPAGLLETNDLASAVFPPLASDEAGNVTLMNSDGLYTKONKL----AVKVGPGLSLDSN 121
QY 54 -PVRIDTOD---SEIKDMVEEHLPLITQOOEVLDEKEQTGFLEAEAPDNVKTMLRSKCY 108
DB 122 NALQVITGDLTVTDKVSMTQAPLSTTSAGLSLLGLPSHLGHEERLTVNT-----174
QY 109 FSKSVSLTEKDAYTHVITPGPRKIANVGVATIGDILDSGNAEYRNALENWQOPVGS 168
DB 175 ---GAGLQTSNNAALAVKVGSGITVDAONQLAASLGDL-----209
QY 169 DFDODSWENSKTSVLG---AVTRKGYPLAKIGNTRAANVPDATTVDLNVVDSGRPIAFG 225
DB 210 -----ESRDNKTVMKAGPLTITNALIVA-TGN-GLQVNP-EQLQLNITAGGLNFANN 262
QY 226 DFEITQRYPEQIVSGLARFOPGT-----PYD-LDL-----LLDPQQAL 264
DB 263 SLAV-----ELGSGL-HFPFGQONVSLYPGGIDIRONRVTVVPAGPLRLMLNQLAV 313
QY 265 EQNGHYSGASVQADRLQGRVPVKVSVT---EVKRHKLETGRLDS-----EYGLGG 315
DB 314 A-----SGDGLVHSDTL---RLKLSHLGFENGAVRAKLGPLGCTDDSGRSVVRTGRGL 365
QY 316 KIAYDYNLFN-KG-YIGSVVMDKYEYTTLLAAGISOPRYNRYNWTSNVSNRSTQML 373
DB 366 RVANGOVQIFSGRGTAGT-----DSSLTLNIRAPLQFSGPALTASLQSGSPITYNS 417
QY 374 EKRAF----SGGIWVVRDRAGIDARLGAEF-----LAEGRKIPGSDIDLG-----NS 416
DB 418 NNGTEGLSIGPQWVDQNRLOVNPAGLIVFGQNLVNPNLADPLAISDSKISLSLPGCLTQ 477
QY 417 HATMLTASWKROLLNNVLPNGHYLDGKIGTITGLTFLSTALRTISARAGYFFTPENKK 476
DB 478 ASNALTL-----LGNLEFSN-QAVALKAGRGU-RPESSQALSSLTAVGNGLT-----525
QY 477 LGTFIIRGOAYTV-ARON---ADVPSGLMFRSGGASVRGYELDSIGLAGPNVSLP-- 530
DB 526 LTTDTVIRPLNGDLEVRDNKIIIVKLANLRFENGAVTA-----GTVNSAPEAPT 576
QY 531 ---BRALLVGSLEYQLPTRLTSLGAVFHD-----MGDA-AANFKRMKKLKHSGSGLGVR 578
DB 577 LTAEPPLRASNSHLQLSL-----EGLVHNNALALQLGDGMVNHQGLTLRLVSGSLQMR 631

RESULT 6
US-09-864-761-37663
Sequence 37663, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmiga-X-1
CURRENT APPLICATION NUMBER: US/09/864.761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/006666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 37663
LENGTH: 794
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005366.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: P33450, EVALUE 5.00e-61
OTHER INFORMATION: EST_HUMAN HIT: BF529515.1, EVALUE 7.00e-78
US-09-864-761-37663

Query Match 3.4%; Score 108.5; DB 10; Length 794;
Best Local Similarity 19.8%; Pred. No. 0.36;
Matches 135; Conservative 75; Mismatches 184; Indels 287; Gaps 35;

QY 25 DUSENKAAG--FALFKNSPDTSYKLPK-----PKFPVPRIDTQDSSEIKDMVEEHLPLI 75
DB 81 ETSEAVAPGTREPLESAHDPOVGSNSLQTYELSRNEYFALRVQTRDSTK-----YAEVLV 135
QY 76 TQOOEEVLDEKQ-----TGFLAEAPDNVKT-----LRSKGYFSKSVS 114

```

Db 136 L---ERADREPERSLQVLTALDGGTPALSASLPPIHKVLDANDNAPVFNOSLYRARVL 192
QY 115 LFEKQAGTVHI-----TPGPRTKI-----ANVGVAIGDILS-----D 148
Db 193 EDAPSGTRVVOVLATDLDEGNGETIYSGSHNRAGVRQLFALDLVTGMLTIKGRLDPEF 252
QY 149 GHLAIVYRNALNNQOPVGS-----DEFODSWENSKTSVLGAVTRKGYPLAKLGNT 199
Db 253 TKLHEIYQAKDKGANPEGACHKVLVEVDVNDNAPETIVTSVSPVED-----APLQTV 308
QY 200 RAAVNPDTATVDLNVVVDGSRPIAFDFEITGTQRYPEQIVSGLARFO--PGTPVDL 257
Db 309 IALL-----SVTDL-----DAGE-----NGLVTCEVPPGLPESLTSS 340
QY 258 LDFQALBONGHYSGASVOADEDRLOGDRVP-----VKSIVTEV 296
Db 341 LK-----NYFLTKTSADLDR---ETVPEYNLSITARDACTPSLSALTIVRQVSDI 388
QY 297 KRHKLETGIRDSEYGLGKAYDY-----NLNKGYIGSVVMDKYEFTLLAAGISOPR 352
Db 389 NDNPQOS-----SOSDYVYTEENLPCAPILNLSVMDPD-----APQ 426
QY 353 NVRGNWTSNVSYNRSTTQNEKRAFSG--GIWTVRDR--AGIDARLGAEEFLAEGRKIPGS 409
Db 427 NARLSFFL-----LEOAGETGLVGRYFTINRDNGIVSSL-----VP-- 462
QY 410 DIDLGNASHATMLTASWKRLQNNVLPENGHYLDGKIGTTLGTFLSSTALITSARAGYF 469
Db 463 -LDYEDRREFELTA-----HISDG-----GTPVLATNISVNIF 494
QY 470 FTEPNKILGTFIIRQAGYTVARDNADVPGLMFRSGA-----SSVRGY 514
Db 495 VTDRN-----DNA--PQVLPYRPGSSVEMLPKRTSAGHLVSRVGVW 534
QY 515 ELD-----SIGLAG--PNGSVL-----PERALLV-----GSLEYOL 543
Db 535 DADAGHNAWLSYLLSGSPNQSFLAIGLHTQISTARPVODTDSPTOTLVLIKDNGEPSL 594
QY 544 PFETRLSGAVFHDMDAAANF 564
Db 595 SITAILTVSVTDSPEARAEF 615

RESULT 7
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-Containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT

```

```

; ORGANISM: streptococcus mutans
US-09-740-274-4
Query Match 3.4%; Score 108.5; DB 10; Length 1375;
Best Local Similarity 19.7%; Pred. No. 0.86;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;

QY 30 KAAGFALFKNKSPTDFSVKLPKFPVVRIDTQSEIKDMVEEH-----LPLITQOQEEV-- 82
Db 729 RPSGVAVIEGNPS-----LRKASDRVVMNMGAAHKNQAYRFLLLITDNGIKA 777
QY 83 --LDKEQTGFL--ABEAPDNVKTMLRSKGYFSKVS-----LFEKDGAVTVHIT 127
Db 778 YHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLCGVWVPVGAADQDVRVAASTA 837
QY 128 PGPRTKIANVGVAIIGDITLSDG-----NLAEYRNA-----LENMOQPVGSDPDQDS 174
Db 838 PSTDGKSVHQNALDSRVMEFGFNFQAFATKKEEYTNVVIKNDVKFENGVTDFEEMAP 897
QY 175 WENSKT--SVLGAVTRKGYPLAKIGNTRAANPDTATVDLNVVVDGSRPIAFG----- 225
Db 898 QYVSSSTDGSLDSVTIONGYAF-----TDYDLGI-----SKPNKYGTADDLVK 940
QY 226 -----DFEITGTQRYPE--QIVSGLARFQPGTP-----YDLDLLLDF 260
Db 941 ATKALHSKGIKVMADWPDQMYALPEREVVATRVKYGTPVAGSQIKNTLIYVD--GK 997
QY 261 QQALEQNGHYSGA-----SVQADFRLQGDVRVPVKSVTEVKRHKLETGIRDSEYGLGCKI 317
Db 998 SSGKDOQAKYGAFFLEELQAKYPEL-----FARKQISTGVMPDPSVKTKOWS 1044
QY 318 A--YDYNNLFNKGYSVVMWM--DKYETTLAAGISOPRN--YRGNWYTSN----- 362
Db 1045 AKYFNGTNILGRG--AGYVYKQOANTYFSLVSDNTFTPLKSLVNPNGHGTSSSTGLVFDGK 1103
QY 363 --VSYNRSTTONLEKRAF--SGGIWYVRDR-----AGIDARLGAEB--FLAEGRKIPGSDI 411
Db 1104 GYVYVY--STSGNQAKNAFISLGNWYTFDNNGYMTVAQOSINGANTYFSLNGTQLRNATY 1161
QY 412 DLGNASHATMLTASWKRLQNNVLPENGHYLDGKIGTTLGTFLSSTALITSARAGYFPT 471
Db 1162 DNGKNVLSYNGDGRY-----ENGYLFGQOWRYFONGIMAVGLTRVHCAVOYFDA 1213
QY 472 PENKKLGTFIIR--GOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLA----- 522
Db 1214 SGFOAKGQFITTADGKLRY--FDRDSGNQISNRFRVN---SKGEMFLFDHNGVAVTGTVTF 1269
QY 523 -----GPNGSVLPERALIV-----GSLEYOLP-----FTRTILSGAVF--HDMGDAA 561
Db 1270 NGORLYFRPNC--VOAKGEFINDANGLYRYDPNSGNEVRNFRVRNKSKEWFLFDHNGTA 1327
QY 562 ANFKRMKLKHGSLGVRWF 580
Db 1328 VTGARVVNGHASILSILMVF 1346

RESULT 8
US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match 3.4%; Score 108.5; DB 10; Length 3158;
Best Local Similarity 20.2%; Pred. No. 3.2;
Matches 91; Conservative 66; Mismatches 159; Indels 135; Gaps 20;

QY 27 SENKAGFALFNKSPDYESVKLPKFPVRIDT-QDSEIKDMVEEHLPLITQQQ----- 79
Db 1004 SLNATGNLI--NATADHQVEQRGNF-INADTDKOTAYTAYNEAEAMINKOTGONANO 1060
QY 80 ---EEVLDKEQTGLAEAPDNVK-----TMLRSKGYFSSKVSLETKDGAYT 123
Db 1061 PEVEQAITTKVQTTLQALNGDNHLOVAKTNATQAI DALTSLSNDPQKTALKDQVTAATLVTA 1120
QY 124 VH-ITPGPRTKANNVGVAILGDLSGHLAEYYRNALENKQOPVGSDFDQDSEKNSKTSV 182
Db 1121 VHQIEQNANT--LNQAMHGLRESIOD--NAATKANSKYINEDQPEQONYDQ-----AV 1169
QY 183 LGAVTRKGYPLAKLGNTRAAYNPDTAVDLNVV-----VDSGRPTAFGDFEITGTQRYCP 236
Db 1170 QAA-----NSIINEQTATLDNNAIQAATVTNTTKAALHGDVKNLQNDKHA 1215
QY 237 EQIVSGLARFQPGTQYDLDLLDFQOALEQNGHYSGASVOADFRLQGDVPPVKVSVTEV 296
Db 1216 KQIVSQALVNNAAQKHMETLIDSETT-----RIAYKQDLTEA 1253
QY 297 K-----RHKLETGI-----RLDSEY---GLGGKIAYDYNNLFNKGIVGYSVWMDMKYET 342
Db 1254 QALDQILMNTLQSIADKDATRASSAYVNAEPNKKQAYD-----EAYONAE 1298
QY 343 TLAAGISOPRNYRGWWSYNSYNRSTTONLEKRAFSGGIWYVRDRAGIDARLGAEFLAE 402
Db 1299 SIAGLNNPTIKGN--VSSATQAVTTSKN-----GLD---GVERLAQ 1336
QY 403 GRKIPGSDIDLGNSHATMLTASWKRQLNNV 433
Db 1337 DKQTAGNSL-----NHLDDQLTQAQOALENQI 1363

RESULT 9

US-09-737-149-25
; Sequence 25, Application US/09737149
; Patent No. US20020077466A1
; GENERAL INFORMATION:
; APPLICANT: Spaderna, Steven K
; APPLICANT: Quinn, Kerry E.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Muralidhara, Padigaru
; APPLICANT: Spyttek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-620 CIP
; CURRENT APPLICATION NUMBER: US/09/737,149

; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/170,564
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: 60/173,165
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,362
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 60/173,544
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/174,404
; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: 60/174,962
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 60/223,929
; PRIOR FILING DATE: 2000-08-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 3034
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-737-149-25

Query Match 3.3%; Score 106; DB 10; Length 3034;
Best Local Similarity 20.3%; Pred. No. 4.9;
Matches 129; Conservative 73; Mismatches 207; Indels 228; Gaps 31;

QY 9 LLPALFFPHAYAPRADLSENKAAGFALFNKSP-DTESVKLPKFPVRIDTDODSEIKDM 67
Db 8 VLPAVLVLLAAALPALEL-----GAAAWELRVPGGARAFALGPGWSYRLDT----- 53
QY 68 VEBHLPLITQOOBEVLDKEBTG-----TRETRELLDVSREGCPAAGRRLGLGAGTLCGLCARLGLRLLPLQVRLVARGCAPTA 98
Db 54 -----TRETRELLDVSREGCPAAGRRLGLGAGTLCGLCARLGLRLLPLQVRLVARGCAPTA 105
QY 99 VKTMLRSKGYFS-SKVSLETKDGATVVIITPGPRTKANNVGVAILGDIL-----SDGNLAE 153
Db 106 PSVLRLARAYGARGCVRLRLRSARGAELRSPAVRS-----VPLGLDALCFPAAGGGA- 158
QY 154 YRNALLENKQOPVGSDFDQDSEKNSKTSVLCGAVTR---KGYPLAKLGNTRAAY--NPDT 207
Db 159 -----SLTSVLEAITNFPACSCPVPVAGTGCNRKGPICLRPG- 193
QY 208 ATVDLNNVVDVSGRPTAFGDFEITGTQRYPEQIVSGLARFQPG-TPYDLDLLDFQOALEQ 266
Db 194 GSLELRLVLCALGRACAGVWVELV-----IQATSGTSPSESPSVSLNLSQPRAGVVR 247
QY 267 NGHYSQASQVQADFRLQGDV-----PVKYSVTEVAKRHKLETGIRLDSEYGLGGKIAYD 320
Db 248 SRRGTGSSSTSPQFP-LPSYQSVSPENEPAGTAVTELRAH-----DPDEGDAGRLSYO 298
QY 321 YYNLFNK---GYI-----GSVV-----WDMKYEITTL-----AAGISOPRNYRGWWSY 363
Db 299 MEALFERSNGYFLIDAATGAVTTARSUDRETQKTHLVKVSADVHDGSPRKSAAATYITVTV 358
QY 364 S-----YNRSTTONLEK-----RAF-----SGGIW--- 383
Db 359 SDTNHSPVFEQSEYRERIRENLEVGEVLITRATOGDAPSNANMRYRLLEGAGGVFEID 418
QY 384 ---YVRDRAGIDARLGAEFLA-----EGRKIPGSDIDLGNSHATMLTASWKRQLNNV 433
Db 419 ARSGVYTRAVYVDRERAEAYQLLVEANDQQRN-PCP-----LSAS---ATVHLIV 463
QY 434 LHPENGHYLDGKIGTGLTGLTFLLSSTALIRTSARAGYFFTPENKKLGTGFIIRQA----- 486
Db 464 VEDENDNY-----PQFSEKRYVQ-----VPEDVAVNTAVLRQVATDRDQGG 505
QY 487 ---GYTVARDNADVPSGIMLPRSGGSSVRYEJDSI 519
Db 506 NAAIHYSIVSGNLKQFYLYHLSLSGSLDVLNPLDFEAI 542

RESULT 10

[illegible]

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5703
; LENGTH: 2025
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(2025)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5703

Query Match 3.2%; Score 102; DB 10; Length 2025;
Best Local Similarity 19.4%; Pred. No. 5.8;
Matches 54; Conservative 49; Mismatches 81; Indels 94; Gaps 11;
QY 176 ENSKTSVLGAVTRKGYPLAKLGNTRAAVNPOTATVDLVNV-----VDSGRPIAFGDFEI 229
DB 29 QNKKTTI-----KVOAANSIINEQTATLDNNAINQAATVNTTKAALHGDVKL 77
QY 230 TGTQRYPEQIVSGLARFOPGTPYDLDLLDFQQALEQNGHYSGASVQADFRLQGDVPV 289
DB 78 QNDKDHAKQTVSQLAVLNNACKHMETLIDSETT-----RIAV 115
QY 290 KVSVTEVK-----RKLETGI-----RLDSEY---GLGKGIADYNNLFNKGITGSVVM 335
DB 116 KODLTEQAQALQDMTLQOSTADKDATRASAYVNAEPNKKQAYD----- 160
QY 336 DMDKYETTLAAGISOPRYGRNYWTSNVSNKSTTONLEKRAFSGGIWYVRDRAGIDARL 395
DB 161 EAVONAESIAGLNNPTINKGN--VSSATQAVTTSKN-----GLD--- 198
QY 396 GAFFLAEGRKIPGSDIDLGNSHATMLTASWKROLLNVV 433
DB 199 GVERLAQDKQTAGNSL----NHLDTLTPAQOQALENOI 232

RESULT 13
US-09-841-132-176
; Sequence 176, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 176
; LENGTH: 982
; TYPE: PRT
; ORGANISM: Chlamydia
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(982)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-841-132-176
Query Match 3.2%; Score 101; DB 10; Length 982;
Best Local Similarity 21.5%; Pred. No. 2.3;
Matches 133; Conservative 70; Mismatches 237; Indels 178; Gaps 32;
QY 108 YFSSKSVLTKDCAYTVHTP---GP---RTKIANVGVAI-----LGDILSD 148
DB 335 FFSSNVAAGGGAIVAKKLSVANCQVQFLRNANDGGAIVLGSGLSLSADYGDIFD 394
QY 149 GNLAEEYRNALENWQOPVGSDFDQDS---WENSKTSVLGAVTRKGYPL-----AKLGN 198
DB 395 GNLK---RTAKENAADVNGVTVSSQAISMGGKITTLLRA--KAGHOILFNDFIEMANGN 449
QY 199 TRAA-----VNP-----DTATVDLVNVVDSGR-----PIAFGDFEITGT 232
DB 450 NQPAQSSKLLKINDGEGYTGDIVFANGSSSTLYQNVTFQGRIVLREKAKLSVNSLSQTGG 509
QY 233 QRYPE-----QIVSGLARFOPGTPYDLDLLDFQQALE-----QNGH- 269
DB 510 SLYMEAGSTLDFVTPPOQPPAANQLITLSNLSLSLLANNNAVNTPTNPPAODSHP 569
QY 270 -----YSGASVQADFRLQGDVPVKSIVTEVKRHKLETG-----IRL 307
DB 570 AVIGSTTAGSVTISGPIFFEDLDLTAYDRYDMLGSNOKINVKLQLGTRPPANAPSDTL 629
QY 308 DSP---YGLGG--KAYDYNNLFNKGYIGSVVWDMDKYETTLAAGISOPRYGRNYWTSN 362
DB 630 GNEPKYIGQGSWKLAOWPNTANNPGYTLKATWTKTGYNPGPERVASLVPN---SLWGS 686
QY 363 V---SYNRSTTONLEKRAFSGGIW-----YVRDRAGIDARLGAEPFLAEGRKIPGSDID 412
DB 687 LDIRSAHSAIQASVDGSRVYCGRLWYSGVSNFFYHDDALGO--GYRIVSGCYSL-GANSY 743
QY 413 LGNSHATMLTASWKROLLNVVLPENGHVLDGKI-----GTTLGLTSLSTALIRTSARAG 467
DB 744 FGSMGLAFTVEFRGSKDYVVCVRNHHACIGSVYLTQQAALCGSYLFGDAFI-----RAS 799
QY 468 YFTTPENKKLGTFIIRGOAGYTVARNADVPFGLMFRSGGASVVRGYELDSIGLAGPNS 527
DB 800 YGFNGOHMK-----TSYTFABE-----SDVR---WDNNCLAGEIGA 832
QY 528 VLP-----ERALLVGSILEQLPRTTLTSGAVPHDM-----GDAAANFKRMKLGKGS-GLGVR 578
DB 833 GLPIVITPSKLYLNELR---FPVQAEFSYADHESFTTEGDOARAFKSGHLLNLSVPVGVK 869
QY 579 W---FSPDLAP--FSPDIAY 592
DB 890 FDRCSSTHPNKYSFMAAY 907
RESULT 14
US-09-841-132-190
; Sequence 190, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay


```
Db 543 VVEHDEDTMASDYLDVPGAGHLGGEIVAAGTPEEVAKNPHSLTGQ--YLSGKKVIPV 600
QY 498 PSGLMFRSGGASSVGYELDSI-----GLAGPNSVLPERALL-----535
Db 601 PKERRKGCKAIKV7GASENNLKNVSEFFLGFBEFVAVTGVSCKSTLVNQILKKAQAK 660
QY 536 -----VGSLEYQLPFTRTLSCAVPHDMGDAAAANFK 565
Db 661 LNRNNSKPKHKHSITGYEABEKIWDQSPIGRTPRSNPATYV---SVFDDIRDLFAQTN 717
QY 566 RMKLKHGSLGVRFSPPLAPSPFIAGCHSDK-----KIRWH 502
Db 718 EAK-----VRGYKK-GRFSFNVKGRCEACRGGDIKIEMH 752

RESULT 16
US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-5639

Query Match 3.2%; Score 100; DB 10; Length 2086;
Best Local Similarity 19.6%; Pred. No. 9;
Matches 103; Conservative 81; Mismatches 215; Indels 126; Gaps 25;

QY 80 EEVLDKEQTGFLEAEAPDN-VKTMRLSKGYFSSKVSUTE-KDGAYTV-----HITPGPRT 132
Db 967 ENILNKGNSNDKAAVENSALQVYTNKAGLNHNLQAKSNANTTINGLOHLTAOKD 1026

QY 133 KIANGVAILGDLSDGCLAEYRYNLENMKNQPPVGSDFDQDSWENSKTSVLGAVTRKGY 192
Db 1027 KL-----KOOVQOQNVAG---VDTVKSSANTLNGA-----1054

QY 193 LAKLGNTRAAVNPDTAT-----VDLNVVDSGRIPAFGDFEITCTORYPEO 238
Db 1055 ---MGLRNSIQDNTATNGONYLDATESKNTYNNAVDS----ANGV1NATSNPNMDAN 1107

QY 239 IVSGIARFQPCPTPYDLDLLDFOQALE-QNGHYSGASVQADEFLRQGDVRPVKVSVEVK 297
```

```
Db 1108 AINOIATQVTSKNAKDGHNLTOAKQATNADGAT---NLNKAOKD--ALKAQVTSQA 1162
QY 298 RIKLETGRLDSEYGLGGKIAYDYNLFNKGYIGSVVWMDKYETTLAAGISQPRNYKGN 357
Db 1163 RVANYSIQOTA-----NELNTA-MGQLQHGIDDENAT-----KOTOKYRDA 1203
QY 358 YWTSNVSY-----NRSTTONLEKRAFSGGIWYV---RDRAGIDARLGAIEFLAEG 403
Db 1204 EOSKRTAYDQAAVAAKATLNKQTCNSDKAAVDRAALQQVYTSKDALNGDAKL-AEAKAA 1262
QY 404 RKPISDDIDLGNSHATMLTASWKRL-----LNVLPENCHYLDGKIGTTLTGLTFLSSTAL 459
Db 1263 KQNLGTLNHIITNAORTALEGOINQATTVDGQNTV--KTNANTLDGAMNSLQGSINDKQAT 1320
QY 460 IRTSARAGYFFTPENKK-----LGTFTIRGQAGYTVARDNADVPSCGLMFRS-----GG 507
Db 1321 LRNQ---NYLDADESKRNAYTQAVTAAGILNKNQTCGNTSKADVDNALWTVTRAKAALNG 1377
QY 508 ASSVRGYELDSIGLAGPNSVLPPE-RALLVGSLEYQLPFTRTLSSG 551
Db 1378 AENLRNTKTSATNTI--NG--LPNLTLQKQDNLKHQVQQAQNVAG 1418

RESULT 17
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match 3.2%; Score 100; DB 10; Length 5795;
Best Local Similarity 19.6%; Pred. No. 45;
Matches 103; Conservative 81; Mismatches 215; Indels 126; Gaps 25;

QY 80 EEVLDKEQTGFLEAEAPDN-VKTMRLSKGYFSSKVSUTE-KDGAYTV-----HITPGPRT 132
Db 3244 ENILNKGNSNDKAAVENSALQVYTNKAGLNHNLQAKSNANTTINGLOHLTAOKD 3303

QY 133 KIANGVAILGDLSDGCLAEYRYNLENMKNQPPVGSDFDQDSWENSKTSVLGAVTRKGY 192
```


Db 3304 KL-----KQVQOQONVAG-----VDTVKSSANTLNGA----- 3331
QY 193 LAKLGNTRAAVNPOTAT-----VDLNVVDSGRPIAFGDFEITGTORYPEQ 238
Db 3332 ---MGLTRNSODTNTATNGQNYLDATESKNNTYNNAVDS-----ANGV1NATSNPNMOAN 3384
QY 239 IVSGLARFOPGPPYDLDLLLOFOQALE--QNGHYSGASVQADFRLQGRVPVKVSVEVK 297
Db 3385 AINQIATQVTSKKNALDCTHNLQAKQATNADGAT--NLNKAOKD--ALKAQVTSAQ 3439
QY 298 RHKLETGRLDSEYGLGKIAYDYNLPNGYIGSVVWDMDKYETTLAAGISQPNRYGN 357
Db 3440 RVANVTSIOOTA-----NELNTA-MGOLOHGIDDENAT-----KOTOKYRDA 3480
QY 358 YWTSNYSY-----NRSTQNLKRAFGSGGIWV---RDRAGIDARLCAEFLEAG 403
Db 3481 EQSKKTAYDOAVAAKAILNKQTCNSDKAAVDRLAQVTSKDALNGDAKL-AEAKAAA 3539
QY 404 RKIPGSDIDLGNSHATMLTASWKQL-----LNNVLHPENCHYLDGKIGTTLGTFSSAL 459
Db 3540 KNLGTLNHIHTNAQRTALEGOINQATTVGVNTV--KTNANTLDGAMNSLQGSINDKQAT 3597
QY 460 IRTSARAGYFFTPENKK-----LGTFIIRGOAGYTVARDNADVPSPGLMFRS-----GG 507
Db 3598 LRNQ---NYLDADESKRNAYTQAVTAAEGILNKQTGGNTSKADVNDNALNTVTRAKAALNG 3654
QY 508 ASSVRGYELDSGLAGPNSVLPD-RALLVGSLEYQLPFTRTLSSG 551
Db 3655 AENLRNTRKTSATNTI--NG--LPNLTLQKDNLKHQVEQAQNVAG 3695

RESULT 18
US-09-897-056-7
; Sequence 7, Application US/09897056
; Patent No. US20020100074A1
; GENERAL INFORMATION:
; APPLICANT: JOE, YUJI
; APPLICANT: MINA, TETSUYA
; APPLICANT: KIDA, TAKAO
; APPLICANT: IGARASHI, DAISUKE
; APPLICANT: OHSUMI, CHIEKO
; TITLE OF INVENTION: METHOD FOR PRODUCING TRANSFORMED PLANT HAVING INCREASED GLUTAMIC
; TITLE OF INVENTION: CONTENT
; FILE REFERENCE: 2104250S0
; CURRENT APPLICATION NUMBER: US/09/897,056
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: JP 2000-215279
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-897-056-7

Query Match 3.1%; Score 99.5; DB 10; Length 1025;
Best Local Similarity 19.7%; Pred. No. 3.3;
Matches 113; Conservative 70; Mismatches 225; Indels 167; Gaps 30;
QY 109 FSSKVSLETKDGAIVHTTPGRTKIANGVVAIIGDILSDGNLAAYRNALENWQOOPVGS 168
Db 41 FHSTILKSRAESA-----APVPRP-----VPLSKLTDGSLDGTSSVILELQRAWAEDPNS 91
QY 169 DFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSGRPIAFGDFE 228
Db 92 --VDESWDNFFRNFVQO-----ASTSPGISQTTQOESMRLL-----LLVRAIQ 132
QY 229 ITG-----TORYPEQIVSGLARFOPGTPYDLDLLLOFOQALEQNGHYSASVQ 276
Db 133 VNGHMKAKLDPLGLEKREIPEDLTPLYGF---TEADLDREF-FLGVWRMSGFLS----- 183
QY 277 ADFDRLQGRDVPVKVSVEVRRHKLKLETKGIRLDSYGLGGKIADYDYNLNFKNKYGISVWVD 336

Db 184 -----ENRPV-----QTLRSILS-----RLQAY--CGTIGYEHIAADKCN---WL 222
QY 337 MKYETTLAAGISOPRNYRGNYWTSNVSYNKSTTONLEKRAFGSGGIWYVRDRAGIDARLG 396
Db 223 ROKIET-----PTPROY--NSERRWVIYDR-LTWSTQPFENFLATKMTAKRFGLE---G 270
QY 397 AEFLAEGRK-----TPGSDIDLGNSHATMLTASWKROLLANNVLH----- 435
Db 271 AESLIPGKMEFDRSADLGVENIVIGHRGL-----NVLGNVVRKRLQIRPSPFSGG 324
QY 436 ----PENGHYL-DGKIGTTLGTFSSALIRTSARAGYFFTPENKKLGTFIIRGQA---G 487
Db 325 TRPVDVGLYTGTDGVKVLHGTSDYRTPRGKHLHLSILVANPNSHLEAVDPVIGKTRAKQ 384
QY 488 YTVARDNADVPSPGLMFRSGGASVVRGYELDSIGLAG-PN-----GSVLPERALLVG----- 537
Db 385 YTKDENRTKNMGIILHGDGSGFAGGVYETLHLSALPNYCTGTV-----HIVVNNQVAF 440
QY 538 ----SLEYOLPFTRTLGSVAFHDMGDAAANFKRMKHLKHSGLGSGVRWFSPLAPSPD 589
Db 441 TTDPREGRSSQCTDVAKALSAPIFHVNADDIE-----AVVHACELAAEW---ROTFFSD 492
QY 590 IA-----YGHSD-----KKIRWHIS 604
Db 493 VVVDLVCYRRRFGHNEIDEPSTQPKMYKVIKSHPS 527

RESULT 19
US-09-747-521-4
; Sequence 4, Application US/09747521
; Patent No. US20020051791A1
; GENERAL INFORMATION:
; APPLICANT: Galloway, Darrel
; APPLICANT: Mateczun, Alfred
; TITLE OF INVENTION: Methods for Protection Against Lethal Infection with Bacillus
; FILE REFERENCE: 22727/04079
; CURRENT APPLICATION NUMBER: US/09/747,521
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-747-521-4

Query Match 3.1%; Score 99; DB 10; Length 764;
Best Local Similarity 18.6%; Pred. No. 2.3;
Matches 111; Conservative 92; Mismatches 221; Indels 172; Gaps 28;
QY 24 ADLSKENKAAGFALFKNKSPTDESVKLP-----KPPVRIDTQDS-ETKQWVEHLPLITQ 77
Db 237 SNIHERK--GLTKYKS-SPEKMWSTASDPYSDFERVTGRIDKNVSPPEARHPLVAAYPVHV 293
QY 78 QQEEVLDDKEQTGFLAEAPDNVKTMLRSKGVFSSKVSLETKDGAYTVHI--TPGPRTKIAN 136
Db 294 DMENIILSKNEODSTONTDSETRTI--SKNTSTRTWTVSEVHGNAEVHANTSTRTWTS 351
QY 137 VGVAILGDLSDGNLAAYRNALENWQOOPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKL 196
Db 352 V-----HGN-AEVHAVAIDHSLSLAG-----ERTWETMGLNTA---DTARL 389
QY 197 GNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFOPGTPYDLDL 256
Db 390 NAMIRVNTGTAPI--YNVLPTTS--LVLGKQNTLATIKAKENQLS----- 431
QY 257 LLDFOQALEQNGHYSASV-----QADFRLQGRDVPVKVSVEVRRHKLKLETKGIRLDS 309
Db 432 -----QILAPNNYVPSKNLPIALNAQDDF-----SSPTIMYNNQPLELETKQLRLDT 481
QY 310 EYGLGCKKIADYDYN-----LFNKYGISG-----VWDM 338

Db 482 DQVYGNATYNFENGRVRYDTSNKNSEVLPOIQTETARIIFNGKDLNLVERRIAANPSD 541
QY 339 KYETT-----LAAGISOPR---NYRGNTWTS-NVSYNRSTTONLEKRAF---SG 380
Db 542 PLETTKPDWTLKALIAFGNPNNGNIQYOGKDTIEFDNFDOOTSONIKNQLABELNAT 601
QY 381 GIWYVRDRAGIDARLGAELFARGRKIPGSDIDL-----NSHATMLTASMKRQLLNIV 433
Db 602 NITVLDKIKLAKNKNILIRDRPHYDRNNIAGVADSVYKAEHREVINSGTEGLLN-- 659
QY 434 LHPENGHYLDGKIGTITLTF-----LSSTALIRTSARAGYFTTPEN 474
Db 660 -----IDKDIRKILSGYIEVEDTEGLKEVINDRYDMLNLISSLRQDGKTFIDFKKN 711
QY 475 KKLGTFTIRQOAG---YTVARNADV-PSGLMFRSGGAS-----SVRGYEL 516
Db 712 DKLPLVISPNKYVNVYAVTKENTINFS-----ENGDTSTNGIKKILIFSCKGYEI 763
RESULT 20
US-10-106-014-4
; Sequence 4, Application US/10106014
; Patent No. US20020142002A1
; GENERAL INFORMATION:
; APPLICANT: Galloway, Darrel R.
; TITLE OF INVENTION: Methods for Protection Against Lethal Infection with Bacillus Anthracis
; FILE REFERENCE: 22727/04114
; CURRENT APPLICATION NUMBER: US/10/106,014
; PRIOR FILING DATE: 2002-03-25
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 764
; TYPE: PR1
; ORGANISM: Bacillus anthracis
US-10-106-014-4

Query Match 3.1%; Score 99; DB 12; Length 764;
Best Local Similarity 18.6%; Pred. No. 2.3;
Matches 111; Conservative 92; Mismatches 221; Indels 172; Gaps 28;
QY 24 ADLSENKAAGFALFNKSPDTESVKLP-----KFPVRIDTQDS-EIKOMVEHPLITO 77
Db 237 SNLHKK--GLTKYKS-SPEKMWSTASDYSDFKVTGRIDKNVSPPEARHPLVAAYPIVHV 293
QY 78 QQEVLDKEQTGFLAEAPDNVKTMLRSKGYFSKVSLETKDGAYTVHI--TPGPRTKIAN 136
Db 294 DMENIILSKNEDQSTQNTDSETRTI--SKNTSTSRHTSEVHGAENVHANTSTSRHTSE 351
QY 137 VGVAILCDILSDGNLAERYRNLWNOOPVCSDFDQDSWENSKTSVLGAVTRKGYPLAKL 196
Db 352 V-----HGN-AEHVAIDAHSLSLAG-----ERTWAETMGLNTA---DTABL 389
QY 197 GNTRAAVNPDTAVLNMVVSGRPIAFDGFELTGRTORYPEQIVSVGLARFQPTPYDLDL 256
Db 390 NANIRVVNTGTAPI-YNVLPTTS--LVLGKNOTLATIKAKENQLS----- 431
QY 257 LLDFOALQNGHYSGASV-----QADPRLOGDRVPVKVSVTEVKRHKLETGIRLDS 309
Db 432 -----DILAPNMYYSKLNAPIALNAODDF-----SSTPTIMYNOFLEKTKQLRLPT 481
QY 310 EYGLGKIAIDYNN-----LFNKYIGSV-----VWMDM 338
Db 482 DQVYGNATYNFENGRVRYDTSNKNSEVLPOIQTETARIIFNGKDLNLVERRIAANPSD 541
QY 339 KYETT-----LAAGISOPR---NYRGNTWTS-NVSYNRSTTONLEKRAF---SG 380
Db 542 PLETTKPDWTLKALIAFGNPNNGNIQYOGKDTIEFDNFDOOTSONIKNQLABELNAT 601
QY 381 GIWYVRDRAGIDARLGAELFARGRKIPGSDIDL-----NSHATMLTASMKRQLLNIV 433

Db 602 NITVLDKIKLAKNKNILIRDRPHYDRNNIAGVADSVYKAEHREVINSGTEGLLN-- 659
QY 434 LHPENGHYLDGKIGTITLTF-----LSSTALIRTSARAGYFTTPEN 474
Db 660 -----IDKDIRKILSGYIEVEDTEGLKEVINDRYDMLNLISSLRQDGKTFIDFKKN 711
QY 475 KKLGTFTIRQOAG---YTVARNADV-PSGLMFRSGGAS-----SVRGYEL 516
Db 712 DKLPLVISPNKYVNVYAVTKENTINFS-----ENGDTSTNGIKKILIFSCKGYEI 763
RESULT 21
US-09-332-226-2
; Sequence 2, Application US/09332226
; Patent No. US20020025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/363,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-332-226-2

Query Match 3.1%; Score 99; DB 10; Length 915;
Best Local Similarity 21.1%; Pred. No. 3;
Matches 145; Conservative 86; Mismatches 270; Indels 186; Gaps 37;

QY 19 AYAPADLSENKAAGFALFNKSPDTESVKLPKFPVRIDTQDSSEIKOMVEHPLITQO 78
Db 22 AYA-----ENVQAQA--QEKQLDTIQVRAKKOKTRR-----DNEVTGLGKLVKVTADTL 69
QY 79 QEEVLD-----KEQTGFLAEAPDNVKTMLRSKGYFSKVSLETKDG-----AYTVHITP 128
Db 70 KEQVLDIIRLTRYDPCIAVYEQGRGASSGYISIRGMKNRVSLT--VDGLAQIQSYTTAAAL 128

```
QY 129 GPRTKIANGVAI-----LGDILSDGNLAERYNALEN---HQCPVGSDF--DODS 174
Db 129 G-GRTAGSGGAINIEIENYKAVKAEISGKSNVSEGGGALAGSAVFOTTTADDVIGEGRQ 187
QY 175 WE-NSKTSVYG---AVTRKGYPLAKLGNTRAAVNPDTATVDLNVVYDSGRPIAFGDFEIT 230
Db 188 WGTQSKTAYSGNRLGTOSIALAGRIGAEA-----LLIRTGRHA--GEIRAH 233
QY 231 GTQRYPEQIVSGLARFORPTPYDLDLLDFOQALEQNGHYSGASVQADFDRLQGDQRPVK 290
Db 234 EAAGRGVSFNRLAPVDGSKYAFIV---EECKNGGH-----EKKAN--PKR 278
QY 291 VSVTEVKRHKLETGRDSEYGLGKIAIDYIN-LFNKG-----YIGSVV---WDM 337
Db 279 DVVGEDKQTVSTRTYTPNRFADPLSYESRSLFRPGRFENKRYHGGILLERTQOTF 338
QY 338 DRYETTLAAGIS-----QPNRYGN-YWTSNVSYNSTTQNLKRA---PSGCI 382
Db 339 DTRDMTVPAFLKAVFDANQKQAGSLRONGKIAGNHKYGGLFTSG-ENNAPVCAEYGTGV 397
QY 383 WYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASWKROLLANNVLPENGHYL 442
Db 398 FY--DETHKSYGLEVY-----VTNADKOTWADYARL---SYDRQ-----GIGL 437
QY 443 DCKICTTGLTFLSSLTALRTSARAGYFTPENKKLGTFIIRG-----QAGYTVARDNAD 496
Db 438 DNHFOQHCSDGSKYCRPSADKPFYSYKSDR-----VIYGESHKLLQAAFKKSPDTAK 492
QY 497 VFSGLMFRSG---GA-----SSVRGYELD-----SIG 520
Db 493 IRHNLVNLGYDRFGSNLRHODYYYSANRAYSLKTPPNNGKKTSPNGREKNPYWVSIG 552
QY 521 LAGPNCVLPERALLVSGLEVLPPTRTLSGAVPHDMDAAAEKRMKLGKSGLGVRWF 580
Db 553 ----RGNVVTRICLFGNANTYDCTPRKINSKSY-----AAVRDNRVLRG-RWA 596
QY 581 SPLAFESDIAYGSHDKKRHWISLGT 607
Db 597 DVGAGLYDRYSTHSD--GSVSTGT 620

RESULT 22
US-09-813-820-6
; Sequence 6, Application US/09813820
; Patent No. US20020102262A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; Pattil, Joseph M.
; House-Pompeo, Karen
; Sthanam, Narayana
; Symersky, Jindrich
; TITLE OF INVENTION: COLLAGEN BINDING PROTEIN COMPOSITIONS
; AND METHODS OF USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/813,820
; FILING DATE: 22-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/856,253
```

```
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: TAMK:193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-813-820-6

Query Match 3.1%; Score 98.5; DB 10; Length 512;
Best Local Similarity 20.5%; Pred. No. 1.3;
Matches 80; Conservative 53; Mismatches 142; Indels 115; Gaps 21;

QY 169 DFDQDSWENSKTSVLGAVTRKGY----PLAKLGNT--RAAVNPDTATVDLNVVYDSGRPI 222
Db 54 DMKIVAMPTS-----GTVKIEGYSKTVPLTVKGBOVGOAVITPDGATITENDKVKLSDV 108
QY 223 A-PGDREITG---TQRYPEQ-----IVSGLARFQPCPTPYDLDLLDFOQALEQNGHYSGA 273
Db 109 SGFAEFVGOGRNLTTQNTSDDKVATITSG-----NKSTNTVYHKSEA 150
QY 274 SVQADFDRLQGDQRPVKVSVTEVK-----RHKLETGRLD----- 308
Db 151 GTSSVYVYKTDMLP--EDTHVRWFLNNEKSYSVSKDITIKDQIGGGQDLSTLIN 208
QY 309 -----SEYGLGKIAIDYINLKNKGYICSVVMDKYEITTLAIGISOPNRYGNWYTSNV 363
Db 209 VITGTHSNYSQGSALTD---FEKAPFGSKI-TVDNTKNTI--DVTIPOGY-GSYNSFSI 260
QY 364 SYNRSSTQNLKRAFSGG-IWY-----VRDRA-----GIDARLGAELAEGR-KIPG 408
Db 261 NYKTKITNEQKEEFVNSQAWYOEHRKEEVNCKSPNHTVHNINANAGIEGTGKGLVK 320
QY 409 SDID-----LGNSHATMLTASWKR-----QLLNNVLPHENGHYLGKGTGT-- 448
Db 321 QDKDTKAPIANVKFKLSKKGDSVVKDNOKEIEITDANGIANIKALPSGDIILKEIAPR 360
QY 449 --TLGTFLSSTALIRTSARAGYFTPENKK 476
Db 381 PYTFDKDKEYPFTMKDTONQGYFTTIENAK 410

RESULT 23
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match
Best Local Similarity 3.18; Score 98; DB 10; Length 6281;
Matches 103; Conservative 83; Mismatches 210; Indels 132; Gaps 25;

Qy 80 EYVDREOTGCFLAEAPON-VKTLMLRSKGYFSSKVSLTE-KGAYTV-----HIVPGPRT 132
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 ENILKNKSGSLDKTAVENALSQVANAKGALNGHNLEQAKSNANTINGLOHLLTAQKD 373

Qy 133 KIANVGVAIIIGDILSDGNLAEYRNALENWOOPVGSDFODSDSHENSKTSVIGAVRTKGY 192
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 374 KL-----KQOVOQAQNVAG-----VDTKVSANTLANGA----- 401

Qy 193 LAKLGNTRAAVNPDTAT-----VDLNVVDSGRPIAFGDFEITGTORYPEQ 238
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 ---MGLTRNSIQDNTATKNGQNYLDATERNKNTYNNAVDS---ANGVINATSNPNMDAN 454

Qy 239 IVSGLARFOPGTPYDLDLLDFQOALE--ONGHYSGSASVQADPRLOGDRVPKVSVEYK 297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 AINOIATQVTSFKNALDQGHNLTKAKOTATNAIDGAT---NLNKAQKD--ALKAOVTSIAO 509

Qy 298 RHILETGIRLDSYGLGKGKIADYVYINFNKGYIGSVVWMDKYETTLAAGISQPNRYGN 357
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 510 RVANVTSIQOTA-----NELNTA-MGQLQHGIDDENAT----KQTKYRDA 550

Qy 358 YFTSNVSY-----NBSTQNLKRAFSGGIHWV---RDRACIDARLCAEFLAEG 403
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 551 EOSKKTAYDQAAVAAKAILNKQTSNDSKAAVDRALQOQVTSKDALNGDAKL-AEAKAAA 609

Qy 404 RKIPGSDIDLGNSHATMLTASWKRL-----LNNYLHPENGHYLDGKIGTIGTFLSTAL 459
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 KQNLGFLNHIHTAQRDLDLEGOINQATTVDGVNTV--KTNANTLDGAMNSLOGSINDKDAT 667

Qy 460 IRTSARAGYEFTPEKK-----LGTFIIRGOAGYTVARNADVPYSGLMFERS----GG 507
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 668 LRNQ---NYLDADESKRNYATQVTAAREGLNKQTKGNTSKADVMDNALNAVTRAKAALNG 724

Qy 508 ASSVRCYELDSI-----GLAGPQSVLPERALLYGSLEYQLPFTRLSG 551
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 725 ADNLRNAKTSATNTIDGL--PNLTQLQK-----DNLKHQVEQAQNVAG 765

RESULT 24
US-09-815-242-13477
; Sequence 13477, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes

```

```
: FILE REFERENCE: ELITRA.011a
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 13477
: LENGTH: 685
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-815-242-13477

Query Match          3.1%; Score 97.5; DB 10; Length 685;
Best Local Similarity 18.5%; Presd. No. 2.6;
Matches 101; Conservative 64; Mismatches 188; Indels 193; Caps 24;

Qy 58 DTQDSBKDWEE--HLPLITQOEEVLD-----KEOTGFLAEAP----- 96
Db 204 DSQAVIASISKEMPGISISWBRKVLETSLSIGSVSSSEKAGLPAEAEAYLKKGYS 263
Qy 97 --DNVKTMLRSKGYSFSSKVSLTEKDGYATVHI-----TGPRTKIANNVV 139
Db 264 LNDRVGTSTYLERQY---BETLOGKRSYKEIHLDKYGNMESVDTIIEEGSKNNIKL-TIDL 319
Qy 140 AILGLDILSDCNLAEYYRNALENWOOPVGSDFDODSWE---NKSQTSV----- 182
Db 320 AFQDSV--DALLKSYFNSELENG----GAKYSGEVYAVALNPKTCAVLMSGCIKHDLKTG 373
Qy 183 -----LGAVTRKGYP-----LAKIGNTRAANVPDTATVDLNWVVDGRPI-----A 223
Db 374 ELTPDSLGGTVTWFPVGSVKAATISSGWNGVLSGNQTLLDQSIVFGSGAPINWSWTQA 433
Qy 224 FGDFEITGTQ-----RYPEQIVSGLAREPQPCTPYDLDLLDDFOALQONGHYSGASVQ 276
Db 434 YGSFPTTAQVALEYSSNTYWVTALGL-----MGOTYQPNFEVCTSNI 477
Qy 277 ADPDRLGQDRVPVKVSYTEVRKHRLTGIRLDSYEGLGGKIADY-----YNLF 325
Db 478 SAMEKLRS-----TFGEYGLGTATGIDLPDESTGFVPKPESYFA 515
Qy 326 NKGYIGSVVMDKYTETTAAGISQPRNYRGNTWTSNVSYN--RSTONLEKRAFPSSGIW 383
Db 516 N--YTNAFPGQDNVTMPQLA-----QYVATIANGVRVAPRIVE-----GIY 556
Qy 384 YVRDRAGIDARLGAEFLEAGRKIPGSDIDILGNSHATMLTASMKROLLNANVLHPENGHYLD 443
Db 557 GNNDKCGGLDILQLOOPTMMKNVWISDSM-----SILHQPFQYVAH 598
Qy 444 KGIGTTLTGFTLSSTALIRTSARACYFTTPENKKLGFTIIRGO-----AGYTVARDNADV 497
Db 599 GTSGLTGTGRAFSNGALVSISGKTG---TAE-----SYVADGQQATNTNAVAYPSDNPQI 650
Qy 498 PSGLMF 503
Db 651 AVAVVF 656

RESULT 25
US-09-815-242-5330
: Sequence 5330, Application US/09815242
```

; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes

RESULT 25
US-09-815-242-5330
; Sequence 5330, A

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5330
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5330
```

```
Query Match 3.1%; Score 97.5; DB 10; Length 801;
Best Local Similarity 20.1%; Pred. No. 3.3;
Matches 95; Conservative 59; Mismatches 199; Indels 119; Gaps 17;
```

```
Qy 25 DLSENKAAGFAL----FNKSPDTEV-----KLKPKFPVRIDTQDSEIKDMVEHLPL 74
Db 109 ELRNSMDAPPVLAYSFSNKENLKKVTKVLNEQLIPKLTQVGVQNAQLNGQTNREITL 168
Qy 75 ITQQEEVLDKEQTGLAEAPDNVKTMLRSKG----YFSKVSLETKDGAYTVHITPGP 130
Db 169 KFKQNE----LEKYGLTADDDVENYLKATRTTPLGLFQFGDKDKSIIVDQY----QSVD 220
Qy 131 RTKIANVGVAILGDILSDGNLAEEYRNALENWQOPVGSDFDQDS--WENSKTSVLGAVTRK 189
Db 221 AFNINIPITLAG--GQGG-----SQSQDNKQNSAMSDVNQASPOQNSKASASNNIS-- 271
Qy 190 GYPLAKIGNTRAANVPDRTATVLDNVVDSGRPIAFGDEITGTORYPEQIVSGLARFPQG 249
Db 272 GNPATKLGQ-----LADITV-----GDVRTSISKTNGKDAVN----- 303
Qy 250 TPYDLDLLDFOQALEONHYSYGASVQADFRLQGRVPVKVSYTEYKRHKLETGIRLDS 309
Db 304 -----LQITKAQDANTVQVAKDVQRKIDTFVDENKDNFNVTMTAKPVEKSL----- 351
Qy 310 EYGLGKGIADYNNFNKGYIGSVV-----WMDKYEITTLAAGISOPNRYGNWYT--- 360
Db 352 -----YTMVERKASLTGTVIAIIVILLFNRNTRTIAISIIPLSLLMALIALKL 399
Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 420
Db 400 SDVSLNLT-----LGAUTVAIGRVDDSIIVVVENIY--RR 433
Qy 421 LTASWKQLLNVLHPENGHYLDGKIGTTLGTFLLSSTALINTSARAGYFTTP 472
Db 434 LTDSERQKGLNLIISATTEVPKPIMSSTLVTIIVFLPLFVSGVSGVGMERP 485
```

RESULT 26

```
US-09-815-242-12189
; Sequence 12189, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12189
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12189
```

```
Query Match 3.1%; Score 97.5; DB 10; Length 1055;
Best Local Similarity 20.1%; Pred. No. 5.1;
Matches 95; Conservative 59; Mismatches 199; Indels 119; Gaps 17;
```

```
Qy 25 DLSENKAAGFAL----FNKSPDTEV-----KLKPKFPVRIDTQDSEIKDMVEHLPL 74
Db 127 ELRNSMDAPPVLAYSFSNKENLKKVTKVLNEQLIPKLTQVGVQNAQLNGQTNREITL 186
Qy 75 ITQQEEVLDKEQTGLAEAPDNVKTMLRSKG----YFSKVSLETKDGAYTVHITPGP 130
Db 187 KFKQNE----LEKYGLTADDDVENYLKATRTTPLGLFQFGDKDKSIIVDQY----QSVD 238
Qy 131 RTKIANVGVAILGDILSDGNLAEEYRNALENWQOPVGSDFDQDS--WENSKTSVLGAVTRK 189
Db 239 AFNINIPITLAG--GQGG-----SQSQDNKQNSAMSDVNQASPOQNSKASASNNIS-- 289
Qy 190 GYPLAKIGNTRAANVPDRTATVLDNVVDSGRPIAFGDEITGTORYPEQIVSGLARFPQG 249
Db 290 GNPATKLGQ-----LADITV-----GDVRTSISKTNGKDAVN----- 321
Qy 250 TPYDLDLLDFOQALEONHYSYGASVQADFRLQGRVPVKVSYTEYKRHKLETGIRLDS 309
Db 322 -----LQITKAQDANTVQVAKDVQRKIDTFVDENKDNFNVTMTAKPVEKSL----- 369
Qy 310 EYGLGKGIADYNNFNKGYIGSVV-----WMDKYEITTLAAGISOPNRYGNWYT--- 360
Db 370 -----YTMVERKASLTGTVIAIIVILLFNRNTRTIAISIIPLSLLMALIALKL 417
Qy 361 SNVSYNRSTTONLEKRAFSGGIWYVDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSHATM 420
Db 418 SDVSLNLT-----LGAUTVAIGRVDDSIIVVVENIY--RR 451
```



```
QY 79 -----QEEVLKKEOTGFL-----AEEA--PD-NVKTMLR-----104
Db 102 EYNQMIEDKRVDFQNNYFKLNNNSKEAIKPDADVSTWKKLDDLDPVMSIFNDFHESPAQ 161
QY 105 -----SKGYFSKYSLTEKGATVHTTGPRTKIANGVAILGIDLSGDNIAEYR 156
Db 162 NEGGOLNGEAWYRKTEFLKDEKDKNNRLTDF-----GVYMSQVYVNGQLVGHYP 213
QY 157 NALENMQPVGSDFDQDQSWNSKTSVLGAVTRKGYPLAKLGNTRAAPNDATVDNLNVV 216
Db 214 NGYNQFSYDITKYLOKQDREN--VIAVHANVKO--PSSRWYSGSGIYRDVTLQVTDKRVH 269
QY 217 D-SGRPIAFGDFEITGTORYPEQIVSGLARFOPCTPYDLDLDDFOQALEQNGHYSGASV 275
Db 270 EKNGTTLTPKLEEQHKGKVEHTVTSKIVNTD---DKDHLEVAEY-QIVERGSHAVTGLV 325
QY 276 QADPDLQODRVPVKVSVTEVKRHKLET-----GTRDSEYGLGSKI 317
Db 326 RTASTLKAHESTSLDALEVERPKLWLVNDKPALYELITRYRDGQLVDKAKDLFG--383
QY 318 AYDYNNLF-NKGYIGSVVWMDKYETTLAAGIS-----QPRNYRGNW-----359
Db 384 -YRYHWTNPNECF--SLNGERIKFH-----GVSLHHDHGALEENYKAEYRRLKOMKEM 435
QY 360 ---TSNVSYNRSTTONLE-----KRAFSGGIWYVRDRAGIDARLG 396
Db 436 GWNSTRTHNPASSOTLOIAELGLLVOEEAFDTWYGGKKPYDGRFFEKDATHPEARKG 495
QY 397 AEFLAEGRKIPGSDIDL-----GNSHATMLTASMKROLNNVLPENCHYLDGK1GTTL 450
Db 496 EKW-----SDFLRTVMVERCKNNPAIF--MWS-----IGNEIGEANGD-----AHS 535
QY 451 GTFLSSTALRTSARAGYFFTPENKKLCTFLIKGOAGYTVARDNADVPGLMFRSGGASS 510
Db 536 ATVKRLVIVKDVKTRT-----LLVGSLEYOLPFTRTLSGAVPHDMGDAANFK 565
QY 511 VRGYELDSIGLAGPNSGYLPERA-----LLVGSLEYOLPFTRTLSGAVPHDMGDAANFK 565
Db 573 IAD-ELDAGVGNYSYEDNYKALRAKHPKWLIIYGETSSATRTR-----GSYYR-----P 619
QY 566 RMLKHGSG 574
Db 620 ERELAHSG 628

RESULT 29
US-09-918-909-24
; Sequence 24, Application US/09918909
; Patent No. US200200907041
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Calmi, Perry G.
; APPLICANT: Orozco Jr., Emil M.
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BB1166 US CIP
; CURRENT APPLICATION NUMBER: US/09/918,909
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 09/697,367
; PRIOR FILING DATE: 2000-10-26
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 1087
; TYPE: PRT
; ORGANISM: Zea mays
US-09-918-909-24
```

Query Match 3.1%; Score 97; DB 10; Length 1087;

```
Best Local Similarity 18.9%; Pred. No. 5.9;
Matches 88; Conservative 63; Mismatches 165; Indels 150; Gaps 19;
QY 16 FPHAYAPAADSENKAAGFALFNKNSPDTSVKLKPKFP-----VRIDQDSEIKD---- 66
Db 157 PEKARQYAADSELDSEGEKGTNNEPSIHDESMRTMRPRIGSTDAIDTWAHQHDKKLY 216
QY 67 --WVEHLPLITQQQEEVLKKEOTGFLAE-----EAPD----- 97
Db 217 IVLIISIHGLIINGENWELGRSDTVKYVVELARALGSLTPGVYRVLDLITROISAPDWDWYSG 276
QY 98 NVKTMRLSKGVSSKVSLETKDGAYTVHTTGPRTKI-----DTATVDNLNVVDSGRPI---- 222
Db 277 EPTMELSPITSENFGLGESSGAYIVRIPFGPRDKYIPKEHLWPHIQEFVGDALVHIMQ 336
QY 135 -----ANVG-----VAILGDILSDGNLAEYRRALENWQOPVGSDFDQDQSWNSKTS 181
Db 337 MSKVLCEQIGSGOPWVPVVIHGHYADAGDSAAALLSGALNVPVPTGHSIGLRK-----389
QY 182 VLGAVTRKGYPLAKLGNTRAAPN-----DTATVDNLNVVDSGRPI---- 222
Db 390 -LDQILKQG-----RQTRDEINATYKIMRRIBAEELCLDTSEI---IITSTRQETEQOW 439
QY 223 -AFGDFEITGTORYPEQIVSGLARF-----OPG-----TPYDLDLDDFOQALE 265
Db 440 GUYDFDFTMARKLRAIRRGVSCFGRYMPRMIAIPGMEFISHIAPHVDL-----DSEE 494
QY 266 QNGHYSGA---SVQADFDR-LQGDVPKVSVTEVKRHK-LETGIRLDSEYG-----312
Db 495 GNGDGGSPDPPTWADIMRFFSNPRKPMILALARPDPKKNITTLVKAFGSHRRLNLANI 554
QY 313 ---LGSKITAYDYNLFNKGYIGSVVWMDKYETTLAAGISQOPNRYR 355
Db 555 TLIENRVDVIDEMSTNAVLTSALKLIDKYD--LYGQVAYPKRHK 598

RESULT 30
US-09-815-242-10728
; Sequence 10728, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10728
; LENGTH: 778
; TYPE: PRT
```

; ORGANISM: Enterococcus faecalis
US-09-815-242-10728

Query Match 3.0%; Score 96.5; DB 10; Length 778;
Best Local Similarity 20.4%; Pred. No. 3.8;
Matches 115; Conservative 67; Mismatches 166; Indels 217; Gaps 31;

QY 83 LDKKQGTGFLA-----EAPDNVY-----TMLRSKGYFSSKVSLETKDGAVT 123
DB 225 LQLPQTALLAGHPQAPNSYDPYTRPTAKERDVVLYTM-----YDNKKLSKAIEYKAKA 279
QY 124 VHTPGPRTKIANVGVAIGDILSDGNIAEYRNALNMQOVPVGSDFDQDSWENSKTSVL 183
DB 280 TPIDEG-----LVPLKASDN-----RKVVONYVYKINEVKARTGKNVYTDGL 323
QY 184 GAVTRKGYPLAKLGNTRAAVNPDTATVDLVN-----VDSGRPIAFGD--FEITGTQRY 235
DB 324 DIYT-----NLDMAKQQLYDIVNSDQYVAFPPDKKQVAST--- 359
QY 236 PEOIVSGLARFQPG---TPYDLDL-----LLDFQOALEBQNGHYSGAS 274
DB 360 VIDVAGQVRAQIGGRHIPDDVQLGNLAVNTORDVGSYVKPIIMDYGPATENLNTSG-- 417
QY 275 VOADFRLQGRVPVKVSVTEVKRHKE-----TGIRLDSYEG--- 312
DB 418 -----RLAVDR-PTKYPGCTDIDVFNSDLYQGVITMRRALMGSRNTTAVOTFDEVGREN 470
QY 313 -----LGGKATDYNNLKNKGYIGSVVWDM--DKY---ETTLAAGISQPRN---YRGNVYT 360
DB 471 IMPFIKG-LGIDYKKNLEASNTSSNTSDVDGDKYGISLKLAAAYAAAFANNGIYKPKYV 529
QY 361 SNVSYNRSYNTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATM 420
DB 530 NKVFNPDGTS-----VDYQPDCKRA-----MKDSTAYM 557
QY 421 LTASWKRQLLNVLPENG-----HYLD-----GKIGTTL-----GTFPL 454
DB 558 MT-----DMLKQVLNCGGTGFNGAIPGLFOAAKTGTSNYTDEDLARMGTTEKGTAPDSTEV 612
QY 455 SSTALIRTSARAGY--FFTP---ENKKGTFPIRQAGY---TVARDNADVPGLMFRSG 506
DB 613 GYTHYAVSVMTGYNDRTPIQVEYGTASDVYREINSYLSQNVSNDDVMVQPSVV-RVG 671
QY 507 GASSVR-GYELDSIGLAGPNSVLP 530
DB 672 NELYVKDAYEV-----PNQVLP 689

RESULT 31
US-09-815-242-5908
; Sequence 5908, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5908
; LENGTH: 1215
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5908

Query Match 3.0%; Score 96.5; DB 10; Length 1215;
Best Local Similarity 19.2%; Pred. No. 7.7;
Matches 85; Conservative 66; Mismatches 173; Indels 119; Gaps 19;
QY 27 SENKAAGFALFNKSPDTSVKLKPFPVRIDT--ODSEIKDMVEEHLPLITQQQ----- 79
DB 500 SLNTANGNLI--NAIADHQAVBQGNF--INADTDKOTAYNTAVNEAAMINKOTQGNANQ 556
QY 80 ---EEVLDEQGTGFLAEAPDNVKTMLRSKGYFSSKVSLETKDGAVTHTTGPRTKI-- 134
DB 557 TEVEQAITKVTQTTLOALNGDHLQV-----AKTNATQATDALT--SLNDPQKTAIKD 606
QY 135 ---ANVGVAILEDILSDGNIAEYRNAL-----ENMQOVPVGSDF---DQDSWENSKTSVL 183
DB 607 QVTAATLVTAHVHQIEQNANTLNQAMHGLRQSIQDNAAATKANSKYINEDQPEQOQYQAVQ 666
QY 184 GAVTRKGYPLAKLGNTRAAVNPDTATVDLVN-----VDSGRPIAFGDPEITGTQRYPE 237
DB 667 AA-----NNIIEQTATLDNNAINQAATVNTTAAALHGDVYKQLNDKDHAK 712
QY 238 QIVSGIARFQPGPYDLDLLDFQOALEBQNGHYSGASVQADEFLRQGRVYVVKVSVTEVK 297
DB 713 QTVSOLAHNLNNAQKHMEDTLIDSETT-----RTAVRQDLTEAQ 750
QY 298 RHKLETGIRLDSY--GLGKATDYNNLKNKGYIGSVVWDMKYE-----TTLAAGISQ 350
DB 751 -----ALDQLMALQOQSIADNRATRASAYVNAEPNKKQSYDEAVQNAESIAGLNN 802
QY 351 PRNYRGNVYTSVSYNRSYNTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSD 410
DB 803 PTINKGNV-----SSATQAV-----ISSKNALD---GVERLAQDKQKTAGNS 840
QY 411 IDLGNSHATMLTASWKRQLLNNV 433
DB 841 L-----NHLQDLTPAQQALENQI 859

RESULT 32
US-09-815-242-13113
; Sequence 13113, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13113
; LENGTH: 1269
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-13113

Query Match 3.0%; Score 96.5; DB 10; Length 1269;
Best Local Similarity 19.2%; Pred. No. 8.3;
Matches 85; Conservative 66; Mismatches 173; Indels 119; Gaps 19;

QY 27 SENRAAGFALPKNSPDTSEKLPKPPVPRIDT-QDSEIKDMVEEHLPLITQQ-----79
DB 554 SLNTANGNLI--NAIADHQAVEQGNF-INADTKOTAYNTAVNEAAAMINKOTGONANO 610
QY 80 ---EEVLDEKGTGLABEAPDNVKTMLRSKGYFSSKVSLEKDGAYTVHIIPGPRKI--134
DB 611 TEVEQAITKVQTTIQAALNGHNLQV-----AKTNATQDAIDALT-SLNDPQKTKALKD 660
QY 135 ---ANVGVAIGDILSDGNLAETRYNAL-----ENQQPVGSDF---DQDSWENSKTSVL 183
DB 661 QVTAATLVTVHQTQANLNTQNMHGLRQSDNATKANSKYINEQPEQNVDAQV 720
QY 184 GAVTRKGYPLAKLGNTRAAVNPDTATVDLNV-----VDSGRPIAFGDFEITGTQRYPE 237
DB 721 AA-----NNIIEQATLDDNAINQAATVNTTAKAALHGDVKLQNDKDHAK 766
QY 238 QIVSGLARFQPGTGYDLDLLDFOALEQNGHYSGASQVADFRLQGDQRPVVKVSVTEVK 297
DB 767 QIVSGLAHLNNAQKHMETLIDSETT-----RTAVKQDLTEAQ 804
QY 298 RHKLETGRLDSEY-GLGKGKIAYDYNNLFPKGYIGSVWMDMKYE-----TTLAAGTSQ 350
DB 805 -----ALDQLMDALQQSADKADATRASAYVNAEPNKKQSYDEAVQNAESIITAGLNN 856
QY 351 PRNYRGNYWTSNVSYNRSTONLEKRAFSGGIWVVRDRAGIDARLGAEFLABGRKIPGSD 410
DB 857 PTINKGV-----SSATQAV-----ISSKNALD---GVERLAQDKQTAGNS 894
QY 411 IDLGNHATMTATSKWKRQLLNIV 433
DB 895 L-----NHLQDLTPAQQAQALENQI 913

RESULT 33
US-09-969-362-3
; Sequence 3, Application US/09969362
; Patent No. US20020076790A1
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schafer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6,-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5340.200-US
; CURRENT APPLICATION NUMBER: US/09/969,362
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 09/397,885
; PRIOR FILING DATE: 1999-09-17

; PRIOR APPLICATION NUMBER: PA 1998 01623
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 60/101,615
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/111,675
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Paenibacillus pabuli
US-09-969-362-3

Query Match 3.0%; Score 96.5; DB 10; Length 1277;
Best Local Similarity 19.4%; Pred. No. 8.3;
Matches 114; Conservative 72; Mismatches 221; Indels 181; Gaps 28;

QY 136 NVGVAILGDILSDGNLAETRYNALLENWQQPVGSDFDQDSWENSK-----TSVLGAVT 187
DB 7 STGCYLLSCVLGSSAAGAYSTASYMETNE--SQMEQTSKESAKOGEAYRMLTKVSKAVT 64
QY 188 R-----KGYPLAK-----LGNTRAAVNPDTATVDLNVV-----DS 218
DB 65 NLSDMTLOGRGSLEDTDEGLHLASEARENVMALSATRA-----DNFIYEDLMIQDKADT 120
QY 219 GRPIAFGDFEITGTQRYPEQIV--SGLARF-----QPGTPYDLDLLDFOALEQNGHY- 270
DB 121 SLVFRSDN---TGWSSYMLQVVPQAGVIRLDASGPGT-----LNVEKHANLESGGIYH 172
QY 271 -----SGASVQADPRLQGDQRPVVKVSVTEVKRHKLETGIRLDSVEYGLG--CKIAYDYN 323
DB 173 LKVKADGESLOVYDNRYPVIDVK-----DSAYSSRGLGLHWDSGA 215
QY 324 LEN-----KGYIGSVWMDMKYETTL-----AAGISQPRN-----YRGNYWT 360
DB 216 LFQNVQVIMGNIGKPISSYGVGEQWPDLAGYKGNAGKQGIYVEKAASDFVEYEGNLYL 275
QY 361 SNVSYNRSTONLEKRAFSGG-----IWVVRDRAGIDARL-----GAEFLEGRKIPGSDI 411
DB 276 ADA-----STSAALLPRASTDGTGKYEAALIREGEEVVRQLRKADGTVLASSNRKVP SOP- 330
QY 412 DLGNHATMTATSKWKRQLLNIVLHPENGHYLDCKITGTTGTLTFLSSTALRTSARAGYF-- 469
DB 331 --GARHHIEVIAS-----GSLIQVYVDGTPAAVEVTDKSYAKGNAGLVVOOGMAYFOD 382
QY 470 -FTPENKKLGTFTIRGOAGYTVARDNADVPGLMFRSG-----GASSVRGYELDSIG 520
DB 383 IYMTESMYKRYENRPPQVHYSPGLGSASDPNGLVYVEGYEYHLPHQDGGTMAHAVSSDLI- 441
QY 521 LAGPNSVLPERRALLVGSLEYQLPFT---RTLGSVPHDMGDAAANF-----564
DB 442 ----NMKRLP-----TALPNDQGHVWGSATADLNNASGLFTDSGGKGLIAYYT 487
QY 565 -----KRMKLKHGSGLGVRW-ESPLAPFSFDIAYGHGSKIRW 601
DB 488 SYHPDKPGNQRIGLAYSTDOGRNWOYAKERPIVIDNPGKNGDDPGSW 535

RESULT 34
US-10-007-693-75
; Sequence 75, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 75

```
: LENGTH: 1609
: TYPE: PRT
: ORGANISM: Chlamydia pneumoniae
US-10-007-693-75

Query Match      3.0%; Score 96.5; DB 12; Length 1609;
Best Local Similarity 19.8%; Pred. No. 12;
Matches 134; Conservative 95; Mismatches 230; Indels 219; Gaps 35;

Qy 69 EEHLPLITQOQEEVLDKEQTGFLAEAEAPDNVKTMLRSGYFSSKVSLETKDG--AVT-V 124
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 49 EHSANVEAQTSLV-----KG--SDVPNSQKESKVLXTQV 84

Qy 125 HITPGPRTKIANGVAILGDIILSDCNLAEYRNALENN-----QQPVGSDFD----- 171
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 PLQGGSSGESLD-----LADANFLEHFQHLFEETVFGIDOKLWSDLTRNFSQP 135

Qy 172 --ODSENKSTSVLGAVTRKGY-----PLAKLG-----NTRAAVNPDTATVDLNVVDS 218
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 TQPDTSNAVSEKISDTHKENRKDLETDPKSKSGLEKVSDDLKSPETAVAAIS----- 190

Qy 219 GRPIAFGDFEITG--TQRYPEQIVGLARFPQPTPYDLDLLDPOQALEONGHY----- 270
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 -----EDLEISENISARPLQ--GLAFFYKNT-----SSQISSEKSDSSFOGIIIFS 233

Qy 271 -----SGASVQAD-----FDRL-----OGDRVPVKVSVT-- 294
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 234 GGSANSGLGFENLAKPKSAAYVSDRDIVFENLVKGLSFSCSELDGSAAGVNIIVTHC 293

Qy 295 -EYKRHKLETG-----IRLDSEYGLGGKI-----AYDYNLFNKGVI-----GSVNDMD 338
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 GDTLTDCATGLDLALRVKDFSRGGAFTARNHIEVQNNLGGILSVVGNKGAIVVEKN 353

Qy 339 KYETTLAAGISOPRNYRGNMTSNVSNRSTTONLEKRAFSGGTYVYDRDRAGIDAR--L 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 SAKSNGGAF-----CGSFVSN--NENTALWKENQALSGGA--ISSASDIDIGNCS 403

Qy 396 GAEF-----LAEGRKIPGSDIDLGNHATMLTASWKR-----QLLNVLHPENGHYLDGK 445
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 AIEFSGNQSLIALGHEHIGITDFVGGGALAAOQTTLRNNAVVQCVRKNTSKTHGGAILAGT 463

Qy 446 IGTTLGTFLSSTALIR--TSARAGYFTFPENKKLGTFTIRGQAGYTVARNADVPGLMFR 504
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 V--DLNETTISEVAFKQNTAALTGGALSANDK-----VIANNFGEILFEQNEVRHGAII 517

Qy 505 SGASSVRGYELDS-----IGLAPNGSV--LPERA-----LLVG 537
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 CCCRSPKLEQKQDSCGNNIILNIGNSGAIITFLKNKASVLEVMTQAEADYAGGGALGHNVLID 577

Qy 538 SLEYQLPFTRTLSSGAVF----HDMGDAANFKRMKLRHSGS-----LGVRFSP 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 SNSGNTQIFGNIIGSTFWIGEYVGGAILSTDRVTISNNSGDVYFKGNKQCLAKRYVAP 637

Qy 583 --LAPTSFDIAYGHSDKK 598
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 638 QETAPVESDASSTNKDEK 655

RESULT 35
US-09-841-132-180
: Sequence 180, Application US/09841132
: Patent No. US20020061848A1
: GENERAL INFORMATION:
: APPLICANT: Bhatia, Ajay
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
: FILE REFERENCE: 210121.469C8
: CURRENT FILING DATE: 2001-04-23
: CURRENT APPLICATION NUMBER: US/09/841.132
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0

: SEQ ID NO 180
: LENGTH: 1752
: TYPE: PRT
: ORGANISM: Chlamydia
US-09-841-132-180

Query Match      3.0%; Score 96.5; DB 10; Length 1752;
Best Local Similarity 20.1%; Pred. No. 14;
Matches 113; Conservative 68; Mismatches 197; Indels 185; Gaps 28;

Qy 108 YFSKSVSLTEKDCAYVHTPGPRTKIANGVAILGD-----ILSDCNLA 152
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1073 FTGNNTATQASSA-----TSGONTNTANYGAAIFGDPGTTQSSQTAULTFLASSGNIT 1127

Qy 153 EYRNALALEMWOQVPWGSDFOODSWENSKTSVLCAVTRKGYPLAKLGNRAAVNPDTATVDL 212
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1128 -FSNNSLQN-----NOGDTPAKSCSIA-----CY-----VKL 1154

Qy 213 NVVVDSCRPIAFODFEIT-----GTORYPEQIVSGLARFPQPTPYDLDLLDPOQALEQ 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1155 SLQAAGKGTLSFDFCVHTSTKKTGTGTONVYETL--DINKENSNPYTGTV--FSSSELHE 1210

Qy 267 NGHY-----SCASVQAD-----FDRLQGDVRVPVKVSVTEVRRHKLETGIRLDSE 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 NKSVIPQNALHNGTLLVLEKTELHVVSFEQKEGSKL-----IMEPCAVLSNQ 1258

Qy 311 YGLGKTAYDYNNLFNKGVI GSVVNDMDKYETTLAAGISOPRNYRGNWTSNVSYNRSTT 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1259 NIANGALA-----INGLTIDLSSMGTPQAGEIFSPPELRIVATTSSA----- 1300

Qy 371 QNLEKRAFSGGIWYVRDRAGIDARLCAEFLEAEGRRKIPGSDIDLGNHATMLTASWKRQLL 430
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1301 -----SGG-----SGVSSSIPTNPKRISAAP-----SGSAATPTMSENKVFL 1339

Qy 431 N--NVLHPENGHYLDGKIGTTLGTFL-----SSTALIRTSARAGYFTFPENKKLGTPI 481
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1340 TGDLTLDIPNGNYQPMILGSDLDVPLIKLPTNTSDVOYDLTJUSGDLF--PQGYMGWT 1398

Qy 482 I--RGQAGYTVARNAD-----VPSGLMFRSGGASSVRGYELDSIGLAPNGSVLPER 532
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1399 LQSNPOTGKLOARWTFDTRVWVYIPRONHFI--ANSILGSQ-----NSMIVVKQ 1446

Qy 533 ALLVGSLEYQLPFTRTLSCAVFHDMDRAANFKRMKLRHSGSLGVRFSPPLAPSPDIAY 592
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1447 GLI-----NNMLNARFDDI--AYNPF-----WVSGVGTFLAOGTFLPSEFSY 1488

Qy 593 ---GHS---DKKIRWHISIGTRF 609
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1489 YSRGTSVAIDAKPRQDFILGAAP 1511

RESULT 36
US-09-815-242-5834
: Sequence 5834, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815.242
: CURRENT FILING DATE: 2001-03-21
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5834
; LENGTH: 2437
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834
```

```
Query Match 3.08; Score 96.5; DB 10; Length 2437;
Best Local Similarity 19.28; Pred. No. 23;
Matches 85; Conservative 66; Mismatches 173; Indels 119; Gaps 19;
```

```
QY 27 SENKAGAFALFNKSPDESVKLPKPPVRIDT-QDSEIKDMVEEHLPLITQQ----- 79
Db 334 SLNTANGNLI--NATADHQAEVQRGNF-INADTKQTAYNTAVNEAAAMINKOTGONANO 390
QY 80 ---EEVLQKEGTGLAEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVHITPGPRTKI-- 134
Db 391 TEVEQAITKVQVTTLOALNGDNRLQV-----AKTNAQTQAIIDALT-SLNDPQRTALKD 440
QY 135 ---ANVGVAIGDILSDGNLAEEYRNAL-----ENWQOPVGSDF---DQDSWENSKTSVL 183
Db 441 QVTAATLVTVHIOIBONANTLNOAHGLOSIODNATKANSKYINEDQPEQONYDQAVQ 500
QY 184 GAVTRKGPYPLAKLGNTRAAVNDPATVDLNVV-----VDSGRPIAFGDFEITGTQRYPE 237
Db 501 AA-----NNHINEQATLDDNAINOAAATVNTTKAALHGDVVKLQNDKDHAK 546
QY 238 QIVSGLARFPQPGTYDLDLDDFQALQEQNGHYSGASVQADFRLQGDVPVKVSVTEVK 297
Db 547 QTVSOLAHNNQAQKHEDTLIDSETT-----RTAVKQDLTEAQ 584
QY 298 RHKLETGIRLSEY-GLGCKIAYIYNLFNKGIGSVVWDMKYE-----TTLAAGISQ 350
Db 585 -----ALDQMDALOOSTADKDAYTRASSAYNAEPNKKQSYDAVQNAESIAGLNN 636
QY 351 PRNYRGNYWTSVSYNRTQNLKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSD 410
Db 637 PTIRKGNV-----SSATQAV-----ISSKNALD---GVERLAQDKQTAGNS 674
QY 411 IDLGNSHATMLTASWKROLLNVV 433
Db 675 L-----NHLDDQLTPAQOQALRNOI 693
```

```
RESULT 37
US-09-815-242-5779
; Sequence 5779, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5779
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5779

Query Match 3.08; Score 95.5; DB 10; Length 841;
Best Local Similarity 17.38; Pred. No. 5.3;
Matches 92; Conservative 77; Mismatches 202; Indels 161; Gaps 22;

QY 92 ABEAPDNVKTMLRSKGYFSSKVSLETKDGAYTVHITPGPRTKIANYGVAILDSDGML 151
Db 79 SRQLKDNTOATADPKVMTSDSATVETSSNMQ---SPQNTATANGSTTKTSNVTNDKS 135
QY 152 ABYYRNALENWQOPVGSDFDQDSWENSKTSVLGAIVTRKGYPLAKLGNTRAAVNDPATVD 211
Db 136 STTYSN-----ETDKSNLTQA-----KDVSTTPKTYI- 163
QY 212 LNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFPQGPY-----DLDDLLDFOQAL 264
Db 164 -----Kp-----RTLNRMVNTVAAPOQGTNVNDKVFHSNIDIAID----- 199
QY 265 EQNGHYSGASVQADFRLQGDVPVKVSVTEVKRHKLETGIRLDSYGLGKIAYDYNNL 324
Db 200 --KGHNQTTGTTEFWATSSDVLKLNKNT-----IDDSVKEGDITTFKYQY 245
QY 325 FNKGIT-----GSV-----VWDMKYE--TLAAGISQPRNRYGNWTSVSY 365
Db 246 FRPGSVRLFSQTNLYNAQGNIIAKGIYDSTTTNTTFTNYVDQYTNVGRSF--EQVAF 303
QY 366 NRSTTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTAS- 424
Db 304 AKRKNATTDKTAYK-----MEVTLGNDTYS-----ELIYDYGKKAQPLISST 347
QY 425 -----WKROLLNVVLPENGHYLDGKIGTTLGTPLSSTALIRT SARAYP----- 469
Db 348 NYINNEDESRNMTAYVNPQKNY-----TKQTFVTNLTGYKFPNPNKFNKFIYEVTDQ 399
QY 470 -----FTPENKKLGTFTIRGQAGYTVARDNADVPGLMFRSGGSGASSVRGKYEIDSLGAG 523
Db 400 NQFVDSFTEDTSKLKD--VTQDFDIYSDNKTATVDLM--KGTSSNKQYLIQQV--AY 453
QY 524 PNGSVLPERALLVGSLEYQLPFTRT---LSCAVFHDMDGDAANAEKRWKLKHG 572
Db 454 PDNSSTDN-----GKIDYTLDTDKTYSWSNSYSNVNGSSSTANGDQKKYNLG 500
```

```
RESULT 38
US-09-815-242-12751
; Sequence 12751, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
```


APPLICANT: Aventis Pasteur Limited
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
FILE REFERENCE: 77813-5
CURRENT APPLICATION NUMBER: US/09/886,468
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 60/113,280
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,281
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,282
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,283
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,284
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,285
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113,385
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114,050
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,056
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,057
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,058
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,059
PRIOR FILING DATE: 1998-12-28
PRIOR APPLICATION NUMBER: 60/114,061
PRIOR FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO. 21
LENGTH: 871
TYPE: PRM
ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match 3.0%; Score 95; DB 10; Length 871;
Best Local Similarity 19.0%; Pred. No. 6.2;
Matches 128; Conservative 84; Mismatches 291; Indels 172; Gaps 28;

QY 5 PTALLPALFFPHAYAPA-ADLS-----ENKAAGFALFKNKSPTDESVKLPKPPVRID 58
DB 6 PWLTTSSALVFSLSHPLMAANTDLSSDNYENGSSGAFTAKETSDAS-----G 54

QY 59 TODSEIKDMVEHLPLITQOOEEVLKDE-----QTGFLAEE--APDNVKT 101
DB 55 TTYTLTSDVITNVSAITPADKSCFTNTGCGALSFVGADHSLVLQTLIALTHDGAAINNTNT 114

QY 102 MLRSKGYFSSKVSUTEKDGAYVHITPGPKTKTANVCVAILGDILSDGNLAEYYRNA--- 158
DB 115 ALSFGSFSLLIDSAPATG-----TSGGKGAICVTN-----TEGGTATFTDNASVT 160

QY 159 -LENNQOPVGSDFDQDQSWENSKSVLCGV-----TRKGYPLAKLGNTRAAVNPDTATVDL 212
DB 161 LQKNTSERKGAASAYSIDLAKTTAALLDQNTSTKNGGALCSTANTTVQGNSTGVTFSS 220

QY 213 NVVDSGRPIAFGPEITGTORYPEQIVSGLARFPQGPY-----DLDLLDFQAL 264
DB 221 NTATDKGGGIYSKEKSDLDAN-----TGVVTFKNTAKTCGAMSSDDNLALTGNTOVL 274

QY 265 EQNGHVSASVQADFDRLQGRVPVKVSVTEVKRHKLETGIRLDSEY----- 311
DB 275 FOENKTTGSAQAANNPECCGAICCYLATATDK-----TGLAISQNOEMSFTSNTTANG 329

QY 312 -----GLGGRKIADYVNLFNKGYIGSVVWMDKYYETTLAA 346
DB 330 GAIYATKCTLDGNTTLTFDQNTATAGCGGAIYETEDFSLKSGTGTVTFTNTAKTGGAL 389

QY 347 GISQPRNRYGNWTNSV--SYNRST---TONLEKRAFSGGINVYVRDRAGIDARLGAEFLLA 401

DB 390 YSKGNSSLTGN---TNLLFSGNKATGPSNSSANQEGCGAILAFIDSGSVSDKTGLS-IA 445
QY 402 EGRKIP-----GSDIDLGNSHATMLTASWKKQLL-----NNVLPENGHY-IDGKIG 447
DB 446 NNQEVSLTSNAATVSGGAIYATKCTLTGNGSLTFDGNATAGTSGGAIYETEDFTLTGSTG 505
QY 448 TTLGTFLSSTALIRTSARAGYFFTPENKKGTFIIRGOAGYTVARDNADVPDG----- 500
DB 506 TV--TFSTNTA-----KTGGALYSKGNHLSL-----GNTNLLFSGNKATGPSNSSANQEG 553
QY 501 -----LMF-RSGGASSVRGY---ELDSIGLAG-----PNGSVLPERALLVG---SLEYOLP 544
DB 554 CGGAILSFLESASVSTKGLWIEDNENVSLSGNTATVSGGAIYATKCALRHGNTTLTFDGN 613
QY 545 FVRTLSGAVFHDMDG 559
DB 614 TAETAGGAIYETED 628

Search completed: November 9, 2002, 01:25:07
Job time : 32 secs

Db 228 NNGYAKAQITKTDVQLNDEKTKVNVTVIDVNEGLOYDLRSARIIGNLGGSAAEPLLSAL 287
QY 150 NLAEEYRNALENNQQPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNDP--- 206
Db 288 HLDNTEFRS-----DIADVENAIKAKLG---ERGY-----GNTTVNSVPDFD 327
QY 207 -TATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFPQGTVPYDLDLLDFOALE 265
Db 328 ANKTLAITFVDAGRRLVROLRFGNTVSADSTLRQEMRQOEGTWYNSOLVELGKIRLD 387
QY 266 QNGHYSAGSQADDFRLOG--DRVPVKVSVTEVRRHKLETGIRLDSYGLGKIAYDYN 323
Db 388 RTGFFE--TVENRIDPINGSNDEVVYKVKERNTGSGINFGIGYGTESGISYQTSIKODN 445
QY 324 LFNKGYIGSVVWMDKYEITTLAAGISOPR-----NYRGNWYTSNVSYNRS-TTONLEKR 376
Db 446 FLGTGAAVSIAGTKNDYGTSVNLGYTEPYFTKDCVSLGGNITFFENYDNSKSDTSSNYKRT 505
QY 377 AFSGGI---WYVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG 414
Db 506 TYGSNVTLGFPPYNNENSYVGLGHTYNNKISNFALEYNRNLYIQSMKFKGNGIKTNDFD-- 563
QY 415 NSHATMLTASWRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFPTPEN 474
Db 564 -----FSGWYNSLNRGYPTKG--VKASLGGGRVTPGSDNKYKLSADVQGFYPLDR 615
QY 475 KKLGTFIIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPN----- 525
Db 616 DHRWVWSAKASAGYANGFNKRLPPYQTYTAGTIGSLRGFPAYGSI---GPNATYAEHNG 672
QY 526 -----GSQLPERALLVGSLEYQLP-----FRT-----L 549
Db 673 TPNKISSDVIIGNAITTASAEILVPTFPVDSKQNTVTSILFVDAASVWNTKWKSDKNGL 732
QY 550 SGAVFIDMGDAANFKRMKLHGSLGVRFPSPLAPSFDA-----YHSDKKIRWHIS 604
Db 733 ESKVLKDLDPYG---KSSRIRASTGCVGQWSPIGPLVFSYAKPIKKYENDDVE-QQFQS 788
QY 605 LGTRF 609
Db 789 IGSF 793

RESULT 2

US-09-135-166-10
; Sequence 10, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-Sep-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-10

Query Match 6.7%; Score 213.5; DB 3; Length 793;
Best Local Similarity 21.0%; Pred. No. 7.8e-11;
Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;

QY 105 SKGYFSSKVSLSL-----EKDGA-VTVHITPGPRTK-----IANVC--VAIIGDILSDG 149
Db 228 NNGYAKAQITKTDVQLNDEKTKVNVTVIDVNEGLOYDLRSARIIGNLGGSAAEPLLSAL 287
QY 150 NLAEEYRNALENNQQPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNDP--- 206
Db 288 HLDNTEFRS-----DIADVENAIKAKLG---ERGY-----GNTTVNSVPDFD 327
QY 207 -TATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFPQGTVPYDLDLLDFOALE 265
Db 328 ANKTLAITFVDAGRRLVROLRFGNTVSADSTLRQEMRQOEGTWYNSOLVELGKIRLD 387
QY 266 QNGHYSAGSQADDFRLOG--DRVPVKVSVTEVRRHKLETGIRLDSYGLGKIAYDYN 323
Db 388 RTGFFE--TVENRIDPINGSNDEVVYKVKERNTGSGINFGIGYGTESGISYQTSIKODN 445
QY 324 LFNKGYIGSVVWMDKYEITTLAAGISOPR-----NYRGNWYTSNVSYNRS-TTONLEKR 376
Db 446 FLGTGAAVSIAGTKNDYGTSVNLGYTEPYFTKDCVSLGGNITFFENYDNSKSDTSSNYKRT 505
QY 377 AFSGGI---WYVRDRAGIDARLG-----AEFLAE-----GRKIPGSDIDLG 414
Db 506 TYGSNVTLGFPPYNNENSYVGLGHTYNNKISNFALEYNRNLYIQSMKFKGNGIKTNDFD-- 563
QY 415 NSHATMLTASWRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFPTPEN 474
Db 564 -----FSGWYNSLNRGYPTKG--VKASLGGGRVTPGSDNKYKLSADVQGFYPLDR 615
QY 475 KKLGTFIIRGQAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPN----- 525
Db 616 DHRWVWSAKASAGYANGFNKRLPPYQTYTAGTIGSLRGFPAYGSI---GPNATYAEHNG 672
QY 526 -----GSQLPERALLVGSLEYQLP-----FRT-----L 549
Db 673 TPNKISSDVIIGNAITTASAEILVPTFPVDSKQNTVTSILFVDAASVWNTKWKSDKNGL 732
QY 550 SGAVFIDMGDAANFKRMKLHGSLGVRFPSPLAPSFDA-----YHSDKKIRWHIS 604
Db 733 ESKVLKDLDPYG---KSSRIRASTGCVGQWSPIGPLVFSYAKPIKKYENDDVE-QQFQS 788
QY 605 LGTRF 609
Db 789 IGSF 793

RESULT 3
US-08-942-046-10
; Sequence 10, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele

QY	475	KKLGTFTIRGOAGYTVARDNADVPVSGLMFSGGASSVVRGYELDSIGLAGPN-----	522
DB	616	DHRVWVSASAGYANGFGNKRLLPFYQTYTAGTGISLRGFAYGSI---GPNATYAEHGMG	672
QY	526	-----GSLPERALLVGSLEYOLP-----FTRT-----L	549
DB	673	TFNKISSDVIGGNAITTASAEELIVTPFVSDKSNQTVRSLFYDAASVHNTKWSKDKNGL	732
QY	550	SGAVFHDGMDAAANFKRMKLUKHSGLGVRWFSPPLAFPSFDIA-----YHSDKKKIRWHIS	604
DB	733	ESKVLKDLDPDYG---KSSRIRASTGVGFQWQSPILGVLVFSYAKPIKKYENDDVE-QPQFS	788
QY	605	IGTRF 609	
DB	789	IGGSF 793	
<p>RESULT 4</p> <p>US-08-433-522A-8</p> <p>: Sequence 8, Application US/08433522A</p> <p>: Patent No. 6013514</p> <p>: GENERAL INFORMATION:</p> <p>: APPLICANT: CHONG, Pelle</p> <p>: APPLICANT: THOMAS, Wayne</p> <p>: APPLICANT: YANG, Yan Ping</p> <p>: APPLICANT: LOOSMORE, Sheena</p> <p>: APPLICANT: SIA, Dwo Yuan Charles</p> <p>: APPLICANT: KLEIN, Michel</p> <p>: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN</p> <p>: NUMBER OF SEQUENCES: 55</p> <p>: CORRESPONDENCE ADDRESS:</p> <p>: ADDRESSEE: Sim & McBurney</p> <p>: STREET: 6TH Floor, 330 University Avenue</p> <p>: CITY: Toronto</p> <p>: STATE: Ontario</p> <p>: COUNTRY: Canada</p> <p>: ZIP: M5G 1R7</p> <p>: COMPUTER READABLE FORM:</p> <p>: MEDIUM TYPE: Floppy disk</p> <p>: COMPUTER: IBM PC compatible</p> <p>: OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>: SOFTWARE: Patentin Release #1.0, Version #1.25</p> <p>: CURRENT APPLICATION DATA:</p> <p>: APPLICATION NUMBER: US/08/433,522A</p> <p>: FILING DATE: 12-SEP-1995</p> <p>: CLASSIFICATION: 435</p> <p>: ATTORNEY/AGENT INFORMATION:</p> <p>: NAME: STEWART, Michael I</p> <p>: REGISTRATION NUMBER: 24,973</p> <p>: REFERENCE/DOCKET NUMBER: 1038-434 MIS:1b</p> <p>: TELECOMMUNICATION INFORMATION:</p> <p>: TELEPHONE: (416) 595-1155</p> <p>: TELEFAX: (416) 595-1163</p> <p>: INFORMATION FOR SEQ ID NO: 8:</p> <p>: SEQUENCE CHARACTERISTICS:</p> <p>: LENGTH: 792 amino acids</p> <p>: TYPE: amino acid</p> <p>: TOPOLOGY: linear</p> <p>: MOLECULE TYPE: protein</p> <p>US-08-433-522A-8</p>			
<p>Query Match</p> <p>Best Local Similarity 6.6%; Score 210.5; DB 3; Length 792;</p> <p>Matches 127; Conservative 85; Mismatches 254; Indels 139; Gaps 24;</p>			
QY	105	SKGYFSKVSFLT-----EKDGA-YTVHITGPRTK-----IANVG--VAILGDLISDG	149
DB	228	NNGYAKAQITKTDVQLNDEKTKVNTIDNGLYDLRSARIIGNLGSMAEPLLSAL	287
QY	150	NLAETRYRNALNNQPVSGDFDQSDWNSKTSVLGAVTRKGYPLAKIGNTRAANVD---	206
DB	288	HLNDFTRS-----DIADVENAIRKALG---ERY-----GNTVNSVPFDD	327
QY	207	-TATVDLNVVDSGRPIAFGFEITGTQRYPEQIVSGLARFPQCTPYDLDLLDFOQALE	265
DB	328	ANKTLAITEVDVAGRLTVRQLRFEQNTVVSADSTLRQEMROQEGTWTYNSQLVELGKTRLD	387
QY	266	QNGHVSQASVQADFRLQG--DRVVPKVSVEVRKHKLETGIRLDSYGLGGKIADVDYN	323
DB	388	RTGFFE--TVENRIDPINGSNDEVVDVYKVERNTGTSINFGIGYGTSGISYQTSIKQDN	445
QY	324	LFNKGYIGSVWMDKQYETTLAAGISQPR-----NVRGNVWTSNVSYNRS--TQNLKLR	376
DB	446	FLGTGAASVIACTKNDYGTSNLGYTEPYFTKDGVSGLGNFFENYDNYSKDSSTSNKRT	505
QY	377	AFSGGT---WYVRDRAGIDARLG-----AFFLAE-----GRKIPGSDIDLGL	414
DB	506	TYGNSVTLGFPVNNNSYVGLGHTYNNKISNFALEYNRNLYIQSKFKNGIKTNDP---	563
QY	415	NSHATMLTASWRQLLNVLHPENGHYLDGKIGTGLTFTLSLTALITSARAGYFFTPEN	474
DB	564	-----FSFGWYNSLNKGYFTTKG--VKASLGRVITPGSDNRYKLSADVOGIFYPLDR	615

```
QY 248 PCTPYDLIDLLFOOALBONGHYSGASVQADFDRLOG--DRVVPKYSVTEVKRHKLETGI 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 EGTWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDDVYKVKERNGTGSINFGI 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 RLDSEYGLGKIAIDYDYNLFNKGYIGSVVWMDKYETTLAAGISQPR-----NYRGNVW 359
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 GYCTESGISYQASVQDNFLGTGAAGVSIAGTKNDYGT'VNLGYTEPYFTKDGVSGLGNVF 487
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 TSNVSNR--TTONLEKRAFSGGI---WYVDRAGIDARLG-----AEFLAE----- 402
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNKSQDTSNTKRTYGSNVTLGFPVNNNSYVVGHTYNNKISNFALEYNNLYI 547
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 403 -----GRKIPGSDIDLGNSHATMLTASWKQLLNVLHPENGYLDGKIGTTLGTFSS 456
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 QSMKFKNGIKTNDFD-----FSGWNYNSLNRGYFTKG--VKASLGGRTVTPGSD 597
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 457 TALIRTSARAGYFTPEPKKLGTFIIRGQAGYTVARONADVPGLMFRSGASSVRGYEL 516
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 NKYYKLSADVQGFPLDRDHLWVYSASAGYANGFCNKRLPFYQYTAGGIGSLRGFAY 657
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 517 DSIGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 GSI---GPNATYQGNKFNKISSDVIGGNAIATASAEILVPTPFVSDKSQNTVRTSLFV 714
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 -----LSGAVFHDMDGAAANFRKMKLKHGSLGLVRFWSPPLAPFSFDIA- 591
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 DAASVWNTKWKSDKNGLSNVLKDLDPYGV---KSRTRASTGVGFQWQSPGVPVFSYAK 771
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 592 ----YCHSDKKIRWHISLGTFR 609
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 PIKKYENDDVE-QFOFSIGGSF 792
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 5
US-09-135-166-8
: Sequence 8, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Sim & McBurney
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1K7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
```

```
TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 792 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-135-166-8

Query Match 5.6%, Score 210.5, DB 3: Length 792;
Best Local Similarity 21.5%, Pred. No. 1.5e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps 21;

QY 134 IANVG--VAIIIGDIIIDGNLAIEYRNALENNQQPVGSDFOQDSWENSKTSLCAVTRKGY 191
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 IGLNGMSAEPLLSALHLNDTPRS-----DIADVENAIRAKLG---ERYG 314
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 PLAKLGNTRAAVNDP----TATVOLNVVDSGRPIACDFEITCTORYEQIVSGLARVQ 247
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 -----GNTTVNSVPDFDDANKTLAITFYVDAGRRLTVIQLRFEGTIVSADSTLRQEMRQ 369
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 248 PCTPYDLIDLLFOOALBONGHYSGASVQADFDRLOG--DRVVPKYSVTEVKRHKLETGI 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 EGTWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDDVYKVKERNGTGSINFGI 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 RLDSEYGLGKIAIDYDYNLFNKGYIGSVVWMDKYETTLAAGISQPR-----NYRGNVW 359
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 GYCTESGISYQASVQDNFLGTGAAGVSIAGTKNDYGT'VNLGYTEPYFTKDGVSGLGNVF 487
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 TSNVSNR--TTONLEKRAFSGGI---WYVDRAGIDARLG-----AEFLAE----- 402
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 FENYDNKSQDTSNTKRTYGSNVTLGFPVNNNSYVVGHTYNNKISNFALEYNNLYI 547
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 403 -----GRKIPGSDIDLGNSHATMLTASWKQLLNVLHPENGYLDGKIGTTLGTFSS 456
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 QSMKFKNGIKTNDFD-----FSGWNYNSLNRGYFTKG--VKASLGGRTVTPGSD 597
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 457 TALIRTSARAGYFTPEPKKLGTFIIRGQAGYTVARONADVPGLMFRSGASSVRGYEL 516
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 598 NKYYKLSADVQGFPLDRDHLWVYSASAGYANGFCNKRLPFYQYTAGGIGSLRGFAY 657
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 517 DSIGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 GSI---GPNATYQGNKFNKISSDVIGGNAIATASAEILVPTPFVSDKSQNTVRTSLFV 714
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 -----LSGAVFHDMDGAAANFRKMKLKHGSLGLVRFWSPPLAPFSFDIA- 591
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 715 DAASVWNTKWKSDKNGLSNVLKDLDPYGV---KSRTRASTGVGFQWQSPGVPVFSYAK 771
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 592 ----YCHSDKKIRWHISLGTFR 609
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 PIKKYENDDVE-QFOFSIGGSF 792
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 6
US-08-942-046-8
: Sequence 8, Application US/08942046
: Patent No. 6264954
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Sim & McBurney
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
```

```

; ZIP: MSG IR7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-8
```

```

Query Match 6.6%; Score 210.5; DB 4; Length 792;
Best Local Similarity 21.5%; Pred. No. 1.5e-10;
Matches 121; Conservative 75; Mismatches 241; Indels 125; Gaps 21;

QY 134 IANVG--VAILGDLSDGNLAERYNALENWQPVGSDPDDDSWNSKTSVLGATRKGY 191
Db 270 IGLGMSAEPLLSALHLNDTFRRS-----DIADVENAIKAKUG---ERYG 314

QY 192 PLAKLGNTRAAVNPD----TATVDLNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFQ 247
Db 315 -----GNTTVNSVPDFFDANKTLAITFVDAGRRLTVHQLRFEGNTVSADSTLRQEMRQ 369

QY 248 PGTPVDLLDLDFQALQNGHYSGASVQADFDFLOG--DRVVPKVSVEVRKHKLEGTGI 305
Db 370 EGTWNSQLVELGKIRLDRDTGFFE--TVENRIDPINSNDEVDVYKVRNFTGSIINFGI 427

QY 306 RLDSEYLGKGTAYDYNYLNFKNYIGSVVMDKYEITTLAAGISQPR-----NYRGNVW 359
Db 428 GYGTSGISYQASVKQDNFLGTGAASVAGTKNDYGTGSVNLGYTEPYFTKDGVSLLGNNVF 487

QY 360 TSNVSYNRS--TTONLEKRAFSGGI--WYVRDRAGIDARLG-----AEFLAE----- 402
Db 488 FENYNSKSDTSNTRKTYGNSVTLGFPVNNNSYVGLGHTYNYKISPALEYNRNLYI 547

QY 403 -----GRKIPGSDIDLGNSHATMLTASWKRQLLNLYLHPENGHYLDGIGTTLGTFLLSS 456
Db 548 QSMKPKGNGIKTNDFD-----FSEGWNYNSLNRGYFPTKG--VKASISLGRVVTIPGSD 597

QY 457 TALIRTSARAGYFETPKENKLGTFILIRGOAGYTVARNADVPISGLMFRSGASSVRGYEL 516
Db 598 NKYYKLASDVQGFYPLDRDLHLWVYSKASAKASYANGFNKRLPYQTYTAGGIGSLRGFAY 657

QY 517 DSTGLAGPN-----GSVLPERALLVGSLEYQLP-----FTRT----- 548
Db 658 GSI--GPNAIYQONKNFKNISISSDVGNGAIAATASAEILIVPTPFVSDKSQNFVTRSLFV 714

QY 549 -----LSCAVFHDMDGAAANFKEMKILKHGSLGVSWFSLAPFSDIA- 591
Db 715 DAASVWNTWKSKNGLNESVULKDLPDYG---KSSRTRASTGVGFQWQSPGVPVPSYAK 771

QY 592 ----YGHSDKKIRWHISLGRF 609
Db 772 PIKKYENDDVE-QQFQSGISGSF 792
```

RESULT 7

```

US-08-433-522A-2
; Sequence 2, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: MSG IR7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-2
```

```

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOBEVLDEQTFGLAEEAPONVKTMRLSKGYFSSKVSLT---- 116
Db 184 ESVSSTLQEOHELOPDSWMLGNKKEGAQFEKDLQSIIRDYLLNNGYAKAQITKTDVOL 243

QY 117 --EKDGA--TVVHITPQPRTK-----IANVG--VAILGDLSDGNLAERYNALENWQOP 165
Db 244 NDEKTRVNTVIDNBEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRS----- 296

QY 166 VGSDFDQDSWNSKTSVLGATRKGYPLAKIGNTRAAVNPD----TATVDLNVVDSGRP 221
Db 297 -----DIADVENAIKAKUG---ERYGSAATNSV-----PDFDANKTLAITLVWDGRR 343

QY 222 IAFGDFEITGTORYPEQIVSGLARFPQTPYDLDLLDFQALQNGHYSGASVQADFDFR 281
Db 344 LTVRQLRFEGNTVSADSTLRQEMRQEGTWTNSQLVELGKIRLDRDTGFFE--TVENRIDP 401

QY 282 LQG--DRVVPKVSVEVRKHKLEGTIRLDSYGLGKGTAYDYNYLNFKNYIGSVVMDMK 339
Db 402 INGSNDEVVYKVRNFTGSIINFGIYGTSGISYQASVKQDNFLGTGAASVAGTKND 461

QY 340 YETTLAAGISQPR-----NYRGNVWTSNVSYNRS--TTONLEKRAFSGGI--WYVRDRA 389
Db 462 YGTSVNLGVTEPYFTKDGVSLLGNNVFEEYNDNSKSDTSNRYKRTYGSNVTLCGFPVNN 521
```

```
Qy 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROL 430
Db 522 SYVVGHTYNNKISNFALEYNRNLYIQSMKPKNGIKTNDFD-----FSGWNYNSL 573
Qy 431 NNVLHPENGHYLDGKIGTTLGTLSTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTV 490
Db 574 NRGYFPTKG--VKASLGGRTVTPGSDNKKYKLSADVOGFPLDRDLHWVYSASASAGYAN 631
Qy 491 ARNADVPGLMFRSGGASSRGVYELDSIGLAGPN-----GSQLPE 531
Db 632 GFNGKRLPFYOTYTAGIGSLRGFAYGSI---GPNAYAEYGNCSGTGTFKKISSDVIGG 688
Qy 532 RALLVGSLEYQLP-----FTRTLSGAVFHDH-----GDAANPKRM-----567
Db 689 NAITATASAEILVPTFPVSDKSONTVRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSGLVGRMPSPLAPSFEDIA-----YGHSDKKIRWHISLGTFR 609
Db 749 KSSRIRASTGCVGFQWQSPIGLPLVFSYAKPIKKYENDDOVE-QFQFSIGGSF 797

RESULT 8
US-08-433-522A-4
; Sequence 4, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433.522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-433-522A-4

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24:
Qy 61 DSEIKDWEEHLPLTIQQEEVLDKEQTFGLAEAPDNKVTMLRSKGYFSSKVSILT----116
Db 184 ESVSSSTLQEMELQSDPMWKLWGNKRFEGAQFEKDLQSIIRDYLLNNGYAKAQITKTDVOL 243
```

```
Qy 117 --BKDGA-YTVHITPCPTK-----IANVG--VAILGDLTSDGNLAEEYVRNALENWQOP 165
Db 244 NDEKTKVNTVDNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRS-----296
Qy 166 VGSDFQDDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPD---TATVDLNVVVDSGRP 221
Db 297 ----DIADVENAIKAKLG---ERCGSATVNSV-----PDFDANKTLAITLWVDAGR 343
Qy 222 IAFGDFEITCTQRYPEQIVYSGLARFPQCTPYDLDLLDFOALEQNGHYVSGASVQADFOR 281
Db 344 LTVRQLRFEGNTVSADSTLRQEMROEGTWTNSQLVELGKIRLDRTOGFFB--TVENRIDP 401
Qy 282 LOG--DRVPPKVSVTEVKRHKLETGIRLDSBYGLGGKIAYDYNNLFNKKYIGSVVWMDMK 339
Db 402 INGSNDDEVVYKVKBRNTGSIKFCIGYCTESGISYQASVKQDNFLCTGAASVIACTKND 461
Qy 340 YETTLAAGISOPR-----NYRCNYWTSNVSYNBS--TONLEKRAFSGGI---WYVRDRA 389
Db 462 YGYSVNLGTEPYFTKDGVSLOGNVFFENYDNSKSDTSSNNYKRTTYGNTVLTGFPVNE 521
Qy 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROL 430
Db 522 SYVVGHTYNNKISNFALEYNRNLYIQSMKPKNGIKTNDFD-----FSGWNYNSL 573
Qy 431 NNVLHPENGHYLDGKIGTTLGTLSTALIRTSARAGYFFTPENKKLGTFIIRGOAGYTV 490
Db 574 NRGYFPTKG--VKASLGGRTVTPGSDNKKYKLSADVOGFPLDRDLHWVYSASASAGYAN 631
Qy 491 ARNADVPGLMFRSGGASSRGVYELDSIGLAGPN-----GSQLPE 531
Db 632 GFNGKRLPFYOTYTAGIGSLRGFAYGSI---GPNAYAEYGNCSGTGTFKKISSDVIGG 688
Qy 532 RALLVGSLEYQLP-----FTRTLSGAVFHDH-----GDAANPKRM-----567
Db 689 NAITATASAEILVPTFPVSDKSONTVRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSGLVGRMPSPLAPSFEDIA-----YGHSDKKIRWHISLGTFR 609
Db 749 KSSRIRASTGCVGFQWQSPIGLPLVFSYAKPIKKYENDDOVE-QFQFSIGGSF 797

RESULT 9
US-08-433-522A-6
; Sequence 6, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433.522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-433-522A-4

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24:
Qy 61 DSEIKDWEEHLPLTIQQEEVLDKEQTFGLAEAPDNKVTMLRSKGYFSSKVSILT----116
Db 184 ESVSSSTLQEMELQSDPMWKLWGNKRFEGAQFEKDLQSIIRDYLLNNGYAKAQITKTDVOL 243
```

REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-433-522A-6

Query Match 6.6%, Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQQOEVLKQETGFLAEAPDNVKTMLRSKGYESSKVSILT--- 116
DB 184 ESVSSSTLOEQMELQPDSSWKLGNKKEGAOFKDLQSIIRDYLLNGYAKAQTITKDVL 243
QY 117 --EKDGA-YTVHITPGPTK-----IANVG--VAIGDLSDGNLAERYRNALENWOOP 165
DB 244 NDEKTKVNTIDVNEGLQYDLRSARIIGNLGMSEALEPLLSALHLNDTFRRS----- 296
QY 166 VGSDFDODSWENSKTSVLGAVTRKGYPLAKLGNTRAANDP---TATVDLNVVDSGRP 221
DB 297 -----DIADVENAIKALG---ERKYGSAVNSV-----PDFDANKTLAITLVVDGRR 343
QY 222 IAFGFEITGTORYPEOIVSGLARFPQGPYDLDLLDFQALQEQNGHYSGASVOADFOR 281
DB 344 LTVQLRFEGNTVSADSTLROEMROOEGTWYNSOLVELKIRLDRTGFFE--TVENRIDP 401
QY 282 LQO--DRVPVKVSVTEVRKHLEGTIRLDSEYGLGKIADYVYVNLNFKYIGSVVWMDK 339
DB 402 INGSNDVYVYKVERNTGSIINGIGYGTESGISYQASVKQDNFLCTGAANSIAGTKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNKSDTSNRYKRTTYGSNVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASMKROLL 430
DB 522 SYVVLGHTYNNKISNFALEYNNLYIOSMKFKNGIKTNDFD-----FSGWYNSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFLSSPTALINTSARAGYFTTPENKLGFTIIRGOAGYTV 490
DB 574 NRGYFPTKG--VKASLGGRTVTPGSDNKYKLSADVOGYPPLDRDHLWVVSASAGYAN 631
QY 491 ARDNADVPSELFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531
DB 632 GFENKRLPYQTYTAGTGLSGRGFAYGSI---GPNAIYAEYNGSGTGCTFKLISSDVIG 688
QY 532 RALLVGSLEYQLP-----FTRLTSGAVFHDH-----GDAAANFKRM----- 567
DB 689 NATATASAEILVPTPEFVSDKSNQTVRTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDY 748
QY 568 ---KLKHGSLGVRWFSPLAPESFDIA-----YCHSDKKIRWHISLGRF 609
DB 749 KSSRIARSTGVGFQWQSPIGPLVSYAKPIKKYENDDVE--QFQFSGISG 797

RESULT 10
US-09-135-166-2
Sequence 2, Application US/09135166
Patent No. 6083743
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,166
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-135-166-2

Query Match 6.6%, Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSEIKDMVEEHLPLITQQOEVLKQETGFLAEAPDNVKTMLRSKGYESSKVSILT--- 116
DB 184 ESVSSSTLOEQMELQPDSSWKLGNKKEGAOFKDLQSIIRDYLLNGYAKAQTITKDVL 243
QY 117 --EKDGA-YTVHITPGPTK-----IANVG--VAIGDLSDGNLAERYRNALENWOOP 165
DB 244 NDEKTKVNTIDVNEGLQYDLRSARIIGNLGMSEALEPLLSALHLNDTFRRS----- 296
QY 166 VGSDFDODSWENSKTSVLGAVTRKGYPLAKLGNTRAANDP---TATVDLNVVDSGRP 221
DB 297 -----DIADVENAIKALG---ERKYGSAVNSV-----PDFDANKTLAITLVVDGRR 343
QY 222 IAFGFEITGTORYPEOIVSGLARFPQGPYDLDLLDFQALQEQNGHYSGASVOADFOR 281
DB 344 LTVQLRFEGNTVSADSTLROEMROOEGTWYNSOLVELKIRLDRTGFFE--TVENRIDP 401
QY 282 LQO--DRVPVKVSVTEVRKHLEGTIRLDSEYGLGKIADYVYVNLNFKYIGSVVWMDK 339
DB 402 INGSNDVYVYKVERNTGSIINGIGYGTESGISYQASVKQDNFLCTGAANSIAGTKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNKSDTSNRYKRTTYGSNVTLGFPVNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASMKROLL 430
DB 522 SYVVLGHTYNNKISNFALEYNNLYIOSMKFKNGIKTNDFD-----FSGWYNSL 573
QY 431 NNVLHPENGHYLDGKIGTTLGTFLSSPTALINTSARAGYFTTPENKLGFTIIRGOAGYTV 490
DB 574 NRGYFPTKG--VKASLGGRTVTPGSDNKYKLSADVOGYPPLDRDHLWVVSASAGYAN 631
QY 491 ARDNADVPSELFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531

```
Db 632 GFCNKRLLPFYQYTAGGSLRGFAYGSI--GPNAIYAEVNGSGTGTFKKLSSDVIGG 688
QY 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFKRM----- 567
Db 689 NAIATASAEILVPTFPVSDKSNQNTVTRSLFVDAASVNTKWKSKNGLESVDLKLRLPDYG 748
QY 568 ---KLKHGSLGVRWFSPFLAPSFEDIA-----YGHSDKKIRWHISLGLTRF 609
Db 749 KSSRIRASTGCVGFQWQSPIGPLVFSYAKPIKKYENDDDVE-QFQFSIGGSF 797

RESULT 11
US-09-135-166-4
; Sequence 4, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jdb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-4

Query Watch 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e+10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDMVEEHLPLITQOOEEYLDKEQTFLEAEAPONVKTMLRSKGFSSKVSILT----116
Db 184 ESVSSSTLQEQMELQPDQSWKWLGNKFKGAQFEKDLQSDIRDYLNNGYAKAQITKTDVOL 243
QY 117 --BKDGA-VTVHITGPRTK-----TANVG--VAILGDTLSDGNLAEEYRNALENWQOP 165
Db 244 NDEKTKVNTIDVNGQLYDLRSARICNLGGMSAELEPLLSALHLNDTFRRS-----296
QY 166 VGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNDP----TATVDLNVVDSGRP 221
```

```
Db 297 -----DIADVENAIRAKLG---ERYGSATVNSV-----PDFDANKTKLTAITLVDAGRR 343
QY 222 IAFCDPEITGCTORYPEQIVSGLARFPQPTPYDLDLDDFOALQEQNGHYSGASVQADFOR 281
Db 344 LTVRQLRFECNTVSADSTLRQEMRQOEGTWNLSQLVELGKIRLDORTGFFE--TVENRNDP 401
QY 282 LOG--DRVPVKYSVTEVKKHKLLETGIRLDSVGLGGKIAIDYYNLFNPKYIGSVVWMDMK 339
Db 402 INGSNDEVDVYKVKERNTGSIINFGIGYGTESGISYQASVKQDNFLGTGAANSIAGTKND 461
QY 340 YETTLAAGISQPR-----NYRCNWTSTNSVYNS--TTONLEKRAFSGGI---WYVRDRA 389
Db 462 YGTSVNLGYTEPYFTKDGVSLLGCVNFFENYDNKSDTSNNYKRTTYGSNNVTLGFPVNNEN 521
QY 390 GIDARLG-----ABFLAE-----GRKIPGSDIDLGNSHATMLTASMKRQLL 430
Db 522 SYVVGLTHTYINKISNFALEYNRLYIQSMKFKGNGIKTNDFD-----FSGWNYNSL 573
QY 431 NNVLHPENGHYLDGKIGITTLGTLSSTALIRTSARAGYFTTPENKKLGLTFLIRGQAGYTV 490
Db 574 NRGYFPTKG--VKASLGRVTTIPGSDNKYYKLSADVQGFYPLDRDHLWVVSASAKASAGYAN 631
QY 491 ARNADVPISGLMFRSGGASSVGVGYELDSIGLAGPN-----GPNAIYAEVNGSGTGTFKKLSSDVIGG 688
Db 632 GFCNKRLLPFYQYTAGGSLRGFAYGSI---GPNAIYAEVNGSGTGTFKKLSSDVIGG 688
QY 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFKRM----- 567
Db 689 NAIATASAEILVPTFPVSDKSNQNTVTRSLFVDAASVNTKWKSKNGLESVDLKLRLPDYG 748
QY 568 ---KLKHGSLGVRWFSPFLAPSFEDIA-----YGHSDKKIRWHISLGLTRF 609
Db 749 KSSRIRASTGCVGFQWQSPIGPLVFSYAKPIKKYENDDDVE-QFQFSIGGSF 797

RESULT 12
US-09-135-166-6
; Sequence 6, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jdb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-6
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-135-166-6

Query Match 6.6%; Score 209.5; DB 3; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDWEHLPLITQOOEEVLKDEQTFGLAEAPDNVMTLRSKGYFSKVSLLT----116
DB 184 ESVSSSTLQOMELQDPSWKLGNKFGAQFEKLOSIRDYILNNGYAKAQTITDVL 243
QY 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAIGDILSDGNLAAYRNALNNWQOP 165
DB 244 NDEKTKVNTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRS-----296
QY 166 VGSDFQDQSWENSKTSVLGAVTRKGYPLAKGNTRAAVNDP-----TATVDLNVVDSGRP 221
DB 297 -----DIADVENAIKAKLG---ERYGTSATVNSV-----PDFDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARFPQGTGYDLDDLLDFQOALEQNGHYSGASVQADFDR 281
DB 344 LTVRQLRFEGNTVSADSTLRQEMRQEGTWTNSQLVELGKIRLDRDTGFFE--TVENRIDP 401
QY 282 LOG--DRVPVKVSVTEVKRHKLETGIRLDSYGLGKTIADYNNLKNKYIGSVVWMDK 339
DB 402 INGSNDVVDVYKVKERNYCSINFGICGTSGISYQASVKQDNFLGTGAANSIATKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTNSVSNYRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNKSDTSSNYKRTYGSNTVLGFPVNENN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROL 430
DB 522 SYVVGHTYTKNISNFALEYNRLYIQSMKFGNGIKTNDPFD-----FSFGWYNLSL 573
QY 431 NNVLHPENGHYLDGKTGTTGLTFLSTALIRTSARAGYFFTPPENKKGITFIIRGQAGYTV 490
DB 574 NRGYFPTKG--VKASLGGRTVTPGSDNRYKLSADVOGFYPLDRDHLWVWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASSVRGYELDSIGLAGPN---GPNAIYAEVNGSGTGTGFKKISSDVIGG 688
DB 632 GFGNKRLLPFQYITAGCIGSLGKPAYGSI---GPNAIYAEVNGSGTGTGFKKISSDVIGG 688
QY 532 RALLVGSLEVQLP-----FTRTISGAVFHDH-----GDAAANPKRM-----567
DB 689 NAITASAEILVTPFVPSDKSONTVRTSLFVDAASVWNTWKSDKNGLSDVLRPLPDYG 748
QY 568 ---KLKHGSLGVNPSPLAPSFDA-----YGHSDKKTRWHSISLCTRF 609
DB 749 KSSIRASTGVGFQWQSPICPLVFSYAKPIKKYENDVE-QFQFSIGCSF 797

RESULT 13
US-08-942-046-2
Sequence 2, Application US/08942046
Patent No. 6254954
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942.046
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/433,522
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-046-2

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

QY 61 DSEIKDWEHLPLITQOOEEVLKDEQTFGLAEAPDNVMTLRSKGYFSKVSLLT----116
DB 184 ESVSSSTLQOMELQDPSWKLGNKFGAQFEKLOSIRDYILNNGYAKAQTITDVL 243
QY 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAIGDILSDGNLAAYRNALNNWQOP 165
DB 244 NDEKTKVNTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRS-----296
QY 166 VGSDFQDQSWENSKTSVLGAVTRKGYPLAKGNTRAAVNDP-----TATVDLNVVDSGRP 221
DB 297 -----DIADVENAIKAKLG---ERYGTSATVNSV-----PDFDANKTLAITLVVDAGR 343
QY 222 IAFGDFEITGTQRYPEQIVSGLARFPQGTGYDLDDLLDFQOALEQNGHYSGASVQADFDR 281
DB 344 LTVRQLRFEGNTVSADSTLRQEMRQEGTWTNSQLVELGKIRLDRDTGFFE--TVENRIDP 401
QY 282 LOG--DRVPVKVSVTEVKRHKLETGIRLDSYGLGKTIADYNNLKNKYIGSVVWMDK 339
DB 402 INGSNDVVDVYKVKERNYCSINFGICGTSGISYQASVKQDNFLGTGAANSIATKND 461
QY 340 YETTLAAGISQPR-----NYRGNWTNSVSNYRS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVSGLGNVFFENYDNKSDTSSNYKRTYGSNTVLGFPVNENN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKROL 430
DB 522 SYVVGHTYTKNISNFALEYNRLYIQSMKFGNGIKTNDPFD-----FSFGWYNLSL 573
QY 431 NNVLHPENGHYLDGKTGTTGLTFLSTALIRTSARAGYFFTPPENKKGITFIIRGQAGYTV 490
DB 574 NRGYFPTKG--VKASLGGRTVTPGSDNRYKLSADVOGFYPLDRDHLWVWSAKASAGYAN 631
QY 491 ARDNADVPSGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531

Db 632 GFGNKRLLPFYQTYTAGGIGSLRGFAVGSII---GPNAIYAEVGNSSGTGTEKKISSDVIGG 688
Qy 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFKRM----- 567
Db 689 NAITASAEELIVPTFPVSDKSONVTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSLGVRWFSPLAFESFDIA-----YHSDKKIRWHISLGRF 609
Db 749 KSSRIRASTGVGFQWQSPICPLVFSYAKPIKKYENDDVE-QQFQSIGGSF 797

RESULT 14
US-08-942-046-4
; Sequence 4, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSHORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 Mis:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-4

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

Qy 61 DSEIKDMVEHLPLITQQEVELDKQETGFLAEAPDNVKTMLRSKGYFSSKVSILT---- 116
Db 184 ESVSSSTLPQMELOPDSWKLWGNKFKGAQPEKDLQSLQSYLLNNGYAKAQITKTDVQL 243
Qy 117 --EKDGA-YTVHITPGPRTK-----IANVG--VAIIGDLSDCNLAEEYRNALENWQOP 165
Db 244 NDEKTKVNTIDVNEGLQYDLRSARIIGNLGMSELEPLLSALHLNDTFERS----- 296
Qy 166 VGSDFDQDSWENSKTSVLGAYTRKGYPLAKIGNTRAANPD-----TATVDLNVVDSGRP 221

Db 297 ----DIADVENAIKAKLG---ERGVGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
Qy 222 IAFGDEFTICTORYPQIVSGLARFQPGTPYDLDLLDLFOALQONQHYSCASVOADFDR 281
Db 344 LTVQRIRFEGNTVSAOSTLRQEMRQOEGTWNSQLVELGKIRLDRDTGFFE--TVENRIDP 401
Qy 282 LQG--DRVPVKVSVTEVKRHKLETGIRLDSYGLGGKTAIDYVNLFNKGYIGSVVMDMK 339
Db 402 INGSNDEVDVVVKYKERNVTSINFGIGYCTESGISYQASVKQDNFLCTGAASVIACTKND 461
Qy 340 YETTLAAGISQPR-----NYRGNTWTSNVSYNRS--TTONLEKRAFSGGI---WYVRDRA 389
Db 462 YGTSVNLGYTEPYFTKDGVSLLGNVFFENYDNSKSDTSSNYKRTTYGSNTVTLGFPVNN 521
Qy 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASWKRQLL 430
Db 522 SYVYVGLGHTYNKISNFALEYNRNLYIQSMKFKNGIKTNDFD-----FSFGWYNSL 573
Qy 431 NVVLHPENGHYLDGKIGTTLGTFLSTALIRTSARAGYFFTPPENKKLGTTIIRGOAGYTV 490
Db 574 NRGYFPTKG--VKASLGGRVTTPGSDNKYKLSADVQGFYPLDRDHLWVVSASAKASYAN 631
Qy 491 ARONADVPSGLMFRSGGASSVRGYELDSIGLAGPN-----GSLVPE 531
Db 632 GFGNKRLLPFYQTYTAGGIGSLRGFAVGSII---GPNAIYAEVGNSSGTGTEKKISSDV 688
Qy 532 RALLVGSLEYQLP-----FTRTLGSAVFHDM-----GDAANFKRM----- 567
Db 689 NAITASAEELIVPTFPVSDKSONVTSLFVDAASVWNTKWKSDKNGLSDVLKRLPDYG 748
Qy 568 ---KLKHGSLGVRWFSPLAFESFDIA-----YHSDKKIRWHISLGRF 609
Db 749 KSSRIRASTGVGFQWQSPICPLVFSYAKPIKKYENDDVE-QQFQSIGGSF 797

RESULT 15
US-08-942-046-6
; Sequence 6, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSHORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 Mis:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-6

TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-942-046-6

Query Match 6.6%; Score 209.5; DB 4; Length 797;
Best Local Similarity 20.0%; Pred. No. 1.9e-10;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;
QY 61 DSEIKDMVEEHLPLITQQOBEVLKQGTGLAEAPDNVKTMLRSKGFSSKVSIT--- 116
DB 184 ESVSSTLQEQMELQPSWMLGNKMGKFEQAQPEKDLQSIQDYLLNNGYAKAKITKTQVL 243
QY 117 --ERDGA-YTHITPGPRTK-----IANVG--VAILGDLSDGNLAEEYNNALENWQOP 165
DB 244 NDEKTKVNTIDVNEGLOYLDRSARIIGNLGGMSAELEPLLSALHNDFFRS----- 296
QY 166 VGSDFQDQSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPD-----TATVDLNVVVDGSRP 221
DB 297 -----DIADVENAIAKILG--ERGYGSATVNSV-----PDFDANKTLAITLVVDAGR 343
QY 222 TAFGDFEITGTORPEOIVSGLARFOPGTPYDLDDLQDQALEQNGHYSGASVOADFDR 281
DB 344 LTVRLREFEGTVSADSTLRQEMRQEGTWYNSQLVELGKIRLDRTGFE--TVENRTDP 401
QY 282 LQG--DRYPVKVSVTEVKRIKLTGRLDSEYGLGKGIAYDYNNLFNKGYIGSVVWMDMK 339
DB 402 INGSNDEVVYKVKERTGSGINFGIGYGTESGISVOASVKODNFLTGTGAASVAGTAKND 461
QY 340 YETTLAAGISOPR-----NYRGNVTSVNSYNS--TTONLEKRAFSGGI---WYVRDRA 389
DB 462 YGTSVNLGYTEPYFTKDGVLGGNVFFENYDNSKSDTSSNYKRTYIGSNVTLGFPVNNN 521
QY 390 GIDARLG-----AEFLAE-----GRKIPGSDIDLGNSHATMLTASMKROLL 430
DB 522 SYVYGLGHTYKINFALEYNRNLVYIOSMKFGKNGIKTNDFD-----FSFGWYNSL 573
QY 431 NNVLHPENGHLVDKIGITGLTFLSSTALIRTSARAGYFTPENKKLGTFTIIRCOAGTV 490
DB 574 NRGYPTKG--VKASLGGRTVPGSNKYKLSADVOGFPYLDRLDHLWVVSASAGAYAN 631
QY 491 ARDNADVPGLMFRSGGASSVRYGLDSIGLAGPN-----GSLVPE 531
DB 632 GPGNKRLPFYQTYTAGGIGSLRGFAYGSI--CPNAIAYEYNGSGTGTFFKKISSDVIGG 688
QY 532 RALLVGSLEYQLP-----FTRLSCAVPHDM-----GDAAANFKRM----- 567
DB 689 NAIATASAEILVPTTFVSDKSQNTVTSLEVDAAVWNTKWSKDRNGLESQVLRKLPDYG 748
QY 568 ---KLKHGSLGVRWFSPAPFSDIA-----YGHSDKKIRWHISLGRF 609
DB 749 KSSRIRASTGVGFQWQSPIGPLVFSVAKPIKKYENDDVE-QFQFSIGSGF 797

RESULT 16
US-09-346-408-6
Sequence 6, Application US/09346408B
Patent No. 6338966
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Anderson, Shawn
APPLICANT: Falco, Carl
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
FILE REFERENCE: BB-1167-A
CURRENT APPLICATION NUMBER: US/09/346,408B
CURRENT FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: 60/092,833

EARLIER FILING DATE: July 14, 1998
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Microsoft Office 97
SEQ ID NO 6
LENGTH: 465
TYPE: PRT
ORGANISM: Glycine max
US-09-346-408-6

Query Match 3.9%; Score 123; DB 4; Length 465;
Best Local Similarity 21.6%; Pred. No. 0.0077;
Matches 108; Conservative 67; Mismatches 183; Indels 142; Gaps 28;
QY 98 NVKTMRLSKGYFSSKVS--LTERDGAITYTHITPG-----PRTKIANVGVAI 141
DB 33 NVEVNFARRRHVGVVRSNALIEPDGKLVVYTDPERDLKKGEALSPLRIKLSRIDL-- 90
QY 142 LGDILSDGNLAEEYNNALENWQOPV-----GSDFDQDQSWNSKTSVLGAVTRKGYPLA-- 194
DB 91 -----EWHVLUSEGNATPLKGFMRAEFTGLHFNLSRLDDGGSVWMSVPIVLA 139
QY 195 -----KLGNTAAVNPDTATVDLNVVVDGSRPIA-FGDFEITGTQYRP--EQIVSGLA 244
DB 140 IDDAQKHRIQD-----NKKVALFD-----SKGDPVAILANNIEI---YKHPRKEERTARTWG 186
QY 245 RFQGTPTP-----DLDLL--LDFQQAQLEQNGHY--SGASVQADFDRLOGD 285
DB 187 TIAPGLTYVEQITNAGNWLIGGDLEVIEPIQYNDGLD---HERLSPAQLRAEFTRRND 243
QY 286 RV-----PV-----KVSVTEVKRHKLETGIR--LDSEYGLGKGIAYDYNNLFNKGYIGS 332
DB 244 AVFAFOLNPNVHNGHALLMTDTRKLEMGYKKNPVLHLLPLGG-----YTKADDDVP 294
QY 333 VVWMDXYETTLAAGISOPRNYRGNVTSVNSYNSRSTTONLEKRAFSGGIWTVVRDRAGID 392
DB 295 LDWRMKQHEKVELEDGVLDPEVTVVSIFPSPMHYAGTEVO-----WHAKAR--IN 342
QY 393 ARIAGAEFLAEGRKIPGSDIDLGNSHAT---MLTASMKROLLNNVLHPENGHYLDGKIGT 448
DB 343 A--GANFYIVGRD-----PAGKSHPVKRDLYADADHKKVLSMAPGLERLNILPFRV-- 392
QY 449 TLGTFLSSTALIRTSARAGYFTPENKKLGTFTIIRCOAGYTVARONADVPSGLMFRSGGA 508
DB 393 -----AAYDKTQGRMA-FEDPSRPQDFLF--ISGTMKMTLARNKSPDPGFMCPGGWK 442

RESULT 17
US-09-346-408-4
Sequence 4, Application US/09346408B
Patent No. 6338966
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Anderson, Shawn
APPLICANT: Falco, Carl
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
FILE REFERENCE: BB-1167-A
CURRENT APPLICATION NUMBER: US/09/346,408B
CURRENT FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: 60/092,833
EARLIER FILING DATE: July 14, 1998
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 465
TYPE: PRT
ORGANISM: Momordica charantia
US-09-346-408-4

Query Match 3.8%; Score 121.5; DB 4; Length 465;

Best Local Similarity 21.7%; Pred. No. 0.011;

Matches 98; Conservative 63; Mismatches 177; Indels 113; Gaps 24;

QY 130 PRTKIANGVAILGDLSDGNLAERYNNALENQOPV-----GSDPDQDSWENSKTSVLG 184

DB 81 PRIELSSIDIQW-HVISEG-----WASPLTGFMRESEFLQTLHFNLSRLPDG 127

QY 185 AVTRKGYPLA-----KLGNTAAVNPDATVDLNVVYDSCRPIA-FCDFEITCTQRYPEQ- 238

DB 128 SVANMSVIVLAIDDAQKRLADSTSTVALFDAN-----NNTAILKDIEI---YKHPEE 179

QY 239 -IVSGLARFPQCTPY-----DLDLL--LDFQQAELQNGHYSGASVQADF 280

DB 180 RIARTWGTAPGLPYVDQAITNAGNWLIGGDLEIVFIKYHVDGLDR-FRQSPAELRESEFT 238

QY 281 RLQGRV-----PV-----KVSYTEVKRHKLETGR--LDSEYGLGKGIAYDYNLENK 327

DB 239 RRDADAFQRLRNPVHNGHALMDTRRLLDGKYNPILLHPLGG-----YTK 289

QY 328 GYIGSVVDMKYEFTLAAGISQPNRYGNWTSNYSNRSTQNLEKRAFSGGIWYVRD 387

DB 290 ADDVPLSRMKQHEKVLGDLDETTVWSIFPSPMHYAGPEVQ-----WHAKA 339

QY 388 RAGIDARLGAELAEGRKIPGSDIDLGNHAT---MLTASMKROLLNVLHPENGHYLD 443

DB 340 R--INA--GANFYIVGRD-----PAGMGHPTKRDLYDADHGKVKLSNAPGLERLNL 389

QY 444 GKGITTLGTFUSSTALITSARAGYFTTPENKKLGTFTIRGQAGYTVARNADVPGLMF 503

DB 390 FRV-----AAYDKTQKMA-FFDPSRPQDFLF-ISGTMRTLAKNKENPPGFM 437

QY 504 RSGGASSVRGYELDSIGLAGPNSVLPERAL 534

DB 438 PGGMKVLVEYD----SLVPSNDRLEPEVL 464

RESULT 18

US-09-346-408-8

; Sequence 8, Application US/09346408B

; Patent No. 6338966

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Anderson, Shawn

; APPLICANT: Falco, Carl

; APPLICANT: Rafalski, Antoni

; FILE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins

; FILE REFERENCE: BB-1167-A

; CURRENT FILING DATE: 1999-07-01

; EARLIER APPLICATION NUMBER: 60/092,833

; EARLIER FILING DATE: July 14, 1998

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 8

; TYPE: PRT

; ORGANISM: Triticum aestivum

US-09-346-408-8

Query Match 3.6%; Score 114; DB 4; Length 461;

Best Local Similarity 22.3%; Pred. No. 0.051;

Matches 109; Conservative 48; Mismatches 195; Indels 136; Gaps 26;

QY 104 RSKGYFSKSVLTERKDGYTHIT--PG-----PRTKIANGVAILGDLSDG 149

DB 38 RAGMSATRSSLIDPDGALVDLVAPPGRSRAALRAEAEALPRVRLAAVDV----- 87

QY 150 NLAERYNNALENQOPVCS-----DFQDQSWENSKTSVLGAYTRKGYPLA----KLGNTR 200

DB 88 ---EWAHVLAEGWASPLRGFMREHYEQCLHFNLSRLPSGGLANMSLPIVLAVDAAKDR 144

QY 201 AAVNPDTATVDLNVVYDSCRPIAFGDFEITGTQR----YP-----EQIVSGLARFQGTPT 252

DB 145 VGRAPDVA-----LAGPDGELLAVLSVEIYPHNKEERLARTNGTTAGLPY 191

QY 253 -----DLDLLDFQQAELQNM---GHY--SGASVQADFRLQGRV----- 287

DB 192 VDEAITPAGNMLIGGDLEVL---QPIKYNDGLDHYRLSPQQLRDEDFRKGADAVFAQL 247

QY 288 --PV-----KVSYTEVKRHKLETGR--LDSEYGLGKGIAYDYNLENKGYIGSVVYMDMK 339

DB 248 RNPVHNGHALMDTRRLLEKGFNPILLHPLGG-----FTKADDVPLPVMEQ 298

QY 340 YTTTLAAGISQPNRYGNWTSNYSNRSTQNLEKRAFSGGIWYVRDAGIDARLGAEL 399

DB 299 HSKVLEDGVLDPETTIVSIFPSPMHYAGPEVQ-----WHAKAR--INA--GANF 344

QY 400 LAEGRKIPGSDIDLGNHATMLTASMKROLLNVLHPENG-HYLDCKIG-TTGLTFLSST 457

DB 345 YIVGRD-----PAGMGHPTKRDLYN---PDHGKVKLSNAPGLERLNLPLKV 389

QY 458 ALIRTSARAGYFTTPENKKLGTFTIRGQAGYTVARNADVPGLMFPSGGSVRGY-EL 516

DB 390 AAYDTVAKKMAFFEPSRQDFLF-ISGTMRTFAKTGENPPDGFMCPCGMKVLVDYVNSL 448

QY 517 DSIGLAGP 524

DB 449 QTEGATAP 456

RESULT 19

US-08-565-386-11

; Sequence 11, Application US/08565386

; Patent No. 5741597

; GENERAL INFORMATION:

; APPLICANT: Bavoil, Patrick M.

; APPLICANT: Haia, Ru-ching

; TITLE OF INVENTION: BACTERIOPHAGE OF CHLAMYDIA PSITTACI

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Roches

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/565,386

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Timian, Susan J.

; REGISTRATION NUMBER: 34,103

; REFERENCE/DOCKET NUMBER: 176/60040

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 716-263-1636

; TELEFAX: 716-263-1600

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 596 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-565-386-11

Query Match 3.6%; Score 113.5; DB 1; Length 596;

Best Local Similarity 23.5%; Pred. No. 0.088;

Matches 107; Conservative 55; Mismatches 214; Indels 79; Gaps 22;

QY 14 FTFPHAYAPADLSENKAAGPALPKNKSPTDESVKLKPFPVRIQTQDSEIKDMVEEHL 73

DB 118 YFVPQVTPDGGYAEENSYDYFGLPTKVANYRH-QVLPLRAYNLFIENYRQENIQSLP 176

QY 74 LITQOQEVLDKQOTGLAE---APDNVTKMLRSK--GYFSSKVSULETEKQAGYVHTP 128
DB 177 VWTGADPKVD-PTTGESEDDAVPYVKLMRNKRKYDYFTSALPGQKQSGVIGITG 235
QY 129 GPRTKIANVGVAI---LGDILSDG-NLAEYIRNLENMWOQPVGSDFDQDSDWENSKTSVLG 184
DB 236 GDSGRPLVHGLAIRSYLSDSDQSFSGVSYVNASQW-----FTAD---GRLTSGM 285
QY 185 AV---TRKGXPL-----AKLGNTRAVNPDATVDLNVVDSGRPIAFGDFEITGTRY 235
DB 286 SVPVGTGPNPDINVPYPSFVGTVAQTGSPSS-----STPFVKGDFPV---Y 332
QY 236 PEQIVSGLARFPQPTPYDLDLLDFQALQONHYSG---ASVQADFRLQGD---RVPV 289
DB 333 VD-----LAASSSVTINSLRNATLQOWFEKSGARYGSRYVESVQGHFGVHLDYRAQRPI 387
QY 290 -----KVSVTEVKRHKLETGIRLSEYGLGKIAY-----DYNNLFNKGYI---GSVVWD 336
DB 388 YLGSKSVSNPVQNS-----STDSVPQNLASAYALSTDTKHLFTKSEVHGFGVIGL 442
QY 337 MD-----KYETTLAAGISQPNRYGNWTSNWSYNRSTTONLEKRAFSGGIWYVRDRAGI 391
DB 443 LSATADLYOQGLROWRSRYSRY-DYVWPTFAHLGEBOPVYNKEIYQSDQDVTMDPSGSAVN 501
QY 392 DARLG-AEFLAEGKIPGSDIDLGNSHATMLTASW 425
DB 502 DVPFGYERYAEYRKPSKYTGLFRSNATGTLDSW 536
RESULT 20
5194600-4
; Patent No. 5194600
; APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;
; HILL, KATHRYN; MEADEN, PHILIP
; TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
; ASSEMBLY AND USE THEREOF
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,316
; FILING DATE: 05-MAR-1990
; SEQ ID NO:4:
; LENGTH: 1365
5194600-4

Query Match 3.6%; Score 113; DB 6; Length 1365;
Best Local Similarity 20.9%; Pred. No. 0.39;
Matches 126; Conservative 91; Mismatches 209; Indels 176; Gaps 34;
QY 41 SPDTESVKLKPKFPVRIDTQDSEIKDMVEHLPLITQOOEVL-----DKEQTGFLAEAP 96
DB 208 SLDCSVEYPLTHPLETLONGSRMSISPOLKKILTYPKELLVGADNKKQLHDEPEEL 267
QY 97 DNVTMLRSKGYFSSKVSLEKDGAVTVHTFG-----PRTKIANYGVAITGLDIL 146
DB 268 RELKLRVTS---LISEFYQYKKDITATLNTKTSIVNNFPLISKLIKIVSSYN---KDII 320
QY 147 S-----DGNLAEYIRNLENMWOQPVGSDFDQDSDWENSKTSVLGAVTRKGYPLAKL 196
DB 321 TSNELNSKGFYDMLGLYING-ONNKITSLTPY-----NLLTALKTEYOSLLKI 369
QY 197 GNTRAAVNPDATVDLNVVDSGRPIAFGDFEITGTRYPEQIVSGLARFGQPGPYDIDL 256
DB 370 TNLQLEPSSKICLDSKFLNK-----FSQFSL-----GKLQNLQP-IMKDLHT 412
QY 257 LLDFOOA-----LEONHYSGA--SVOADFRLQGDVRPVKVSYTEVKKR-----299
DB 413 IPGFSSEVIYFNIESDPQDELVNSVQAFDQKSPGELP-----EIKQNWSEIIFVID 466
QY 300 --KLETGIRLSEY--GLGKIAYDYNLFNKGY---IGSVVWMDKTYETTLAAGISQPR 352
DB 467 FARLE-----DSEVKALGLGLVR--AVNVVSGYQPVORVGLLPLFPSSDSKSVY-----511

QY 353 NYRGNWTSNWSYN-RSTTONL-EKRAFSGGIWYVRDRAGIDARLGAEFLABGRK----I 406
DB 512 -----NKIYELKKNSTONLTELKSF-----LETMLLIADGLSANAKHSHI 550
QY 407 PGSDI-----DLGNSHATMLTASWKROLLNNVLHP--ENGHYLDGKIGTTLGTFLSSTAL 459
DB 551 PVPDVEHLDELQIDETSI-----IINGEYTPFRKNWNYLIAKV-----IKKDTF 596
QY 460 IRTSARAGYFTTPENKKLCTFIIRGQAGTYVA--RONADVPSGLMFRSGGASSVGRGYELD 517
DB 597 IRKELNSG---SPANKOIS---VRDLLHYKSANLRHKNKYTPN--YFADSVSYSVNNTALE 648
QY 518 SIGLAGPNGSYLPYR-ALLVGSLEYQLPFTTLLSGAVFHDMDGDAANFRMKLKHGSGIG 576
DB 649 -----SVCSEIRGYTKNEEYNLLHTITL---VDFSGSIHALKRLRLNLLHTSEVG 695
QY 577 VR 578
DB 696 VR 697
RESULT 21
US-08-021-601-10
; Sequence 10, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-10

Query Match 3.5%; Score 112.5; DB 1; Length 508;
Best Local Similarity 20.5%; Pred. No. 0.083;
Matches 112; Conservative 78; Mismatches 189; Indels 167; Gaps 28;
QY 38 KNSPDTSVSLKPKFPVRIDTQDSEIKDMVEHLPLITQOOEVLDKQOTGFLAEAPD 97
DB 14 KEKNKEN-----KRKDERNKQTQEHLEKIM-KHIVKIEVKGEAEVKAEEKLEKVP 68

```
QY 98 NVKTLMSKGYFSSKVLTEKDGAYTVHIT-----PGPRTKIANVGVAIILGDLSDGNLAE 153
Db 69 DVLENYKAIG---GKIYIV---DGDITKHISLEALSDEKKKIKD-----IYG---KDALHE 116
QY 154 YRNALENWQPVGSDFDQDSDWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLN 213
Db 117 HYVYAKEGEPVLVIQSSDYVENTEKA-----LN 146
QY 214 VVDSGRPIAFDPEITGTORYPEQIVSGLARFPQGPYD--LDLLDFQQALEONGHYS 271
Db 147 VYIEIGTKILS-----RDILSKI-----NQYQKFLDVLNTIKNASDSGQ-- 186
QY 272 GASVQADFDRLQGDQRPVKVSVTEVKRHKLETGIRLDSYGLGGKIAYDYNNLFNKGYIG 331
Db 187 -----DLL-----FTNQLKEHPTDFSVEFLEQN-----SNEVOEVFAKAF-- 221
QY 332 SYVWDMKXYETTLAAGISQPRNYRGNYWTSNVSYNSTTONLEK-----RAFS 379
Db 222 -AYIEPQHRDVLQLYAPEAFNYMDKFNEQEIINLTRAANADVVSUTCPVAEGECAGPADS 280
QY 380 GGIVYVRDRAGIDARL-----GAEFLEAGRKIPGSDIDLGNSHATMLTASW--KRQLLN 431
Db 281 G-----DALLERNYPTGAFLGDDGDV-----SFSTRGTQNTWTVLERLLOA 320
QY 432 NVLHPENHYLDGKIGTGTFLSSLTALITSARACYFTTPENKKLGT----FIIRGO-- 485
Db 321 HRQLEERGIVFVGYHGTFLE---AAQSIVFEGVRA-----RSQDLDAIWRGFIAGDPA 371
QY 486 AGYTVARDNADVPGLMFRSGG-----ASSVRGYELDSIGLAGPNSVLPERALLVGS 538
Db 372 LAYGVAQDOEPDARG-RIRNGALLRVVPRSSLPGFYRTSLTLAAPEAAAGEVER--LIG- 427
QY 539 LEYQLP 544
Db 428 --HPLP 431

RESULT 23
US-08-082-849B-10
; Sequence 10, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082.849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-10

Query Match 3.5%; Score 112.5; DB 1; Length 508;
Best Local Similarity 20.5%; Pred. No. 0.083;
Matches 112; Conservative 78; Mismatches 189; Indels 167; Gaps 28;

QY 38 NKSPDTSVKLKEKFPVRIOTDSEIKDMVEEHLPLITQOOEEVLDKDEOTGFLAEAPD 97
Db 14 KEKNKEN-----KRKDERNKTOEHLKEIM-KHIVKEVKEEAIVKRAEAKLCKVPS 68
QY 98 NVKTLMSKGYFSSKVLTEKDGAYTVHIT-----PGPRTKIANVGVAIILGDLSDGNLAE 153
Db 69 DVLENYKAIG---GKIYIV---DGDITKHISLEALSDEKKKIKD-----IYG---KDALHE 116
QY 154 YRNALENWQPVGSDFDQDSDWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLN 213
Db 117 HYVYAKEGEPVLVIQSSDYVENTEKA-----LN 146
QY 214 VVDSGRPIAFDPEITGTORYPEQIVSGLARFPQGPYD--LDLLDFQQALEONGHYS 271
Db 147 VYIEIGTKILS-----RDILSKI-----NQYQKFLDVLNTIKNASDSGQ-- 186
QY 272 GASVQADFDRLQGDQRPVKVSVTEVKRHKLETGIRLDSYGLGGKIAYDYNNLFNKGYIG 331
Db 187 -----DLL-----FTNQLKEHPTDFSVEFLEQN-----SNEVOEVFAKAF-- 221
QY 332 SYVWDMKXYETTLAAGISQPRNYRGNYWTSNVSYNSTTONLEK-----RAFS 379
Db 222 -AYIEPQHRDVLQLYAPEAFNYMDKFNEQEIINLTRAANADVVSUTCPVAEGECAGPADS 280
QY 380 GGIVYVRDRAGIDARL-----GAEFLEAGRKIPGSDIDLGNSHATMLTASW--KRQLLN 431
Db 281 G-----DALLERNYPTGAFLGDDGDV-----SFSTRGTQNTWTVLERLLOA 320
QY 432 NVLHPENHYLDGKIGTGTFLSSLTALITSARACYFTTPENKKLGT----FIIRGO-- 485
Db 321 HRQLEERGIVFVGYHGTFLE---AAQSIVFEGVRA-----RSQDLDAIWRGFIAGDPA 371
QY 486 AGYTVARDNADVPGLMFRSGG-----ASSVRGYELDSIGLAGPNSVLPERALLVGS 538
Db 372 LAYGVAQDOEPDARG-RIRNGALLRVVPRSSLPGFYRTSLTLAAPEAAAGEVER--LIG- 427
QY 539 LEYQLP 544
Db 428 --HPLP 431

RESULT 23
PCT-US94-01624-10
; Sequence 10, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
; STREET: Steuart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
```

STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-10

Query Match 3.5%; Score 112.5; DB 5; Length 508;
Best Local Similarity 20.5%; Pred. No. 0.083;
Matches 112; Conservative 78; Mismatches 189; Indels 167; Gaps 28;
QY 38 KKKSPDSESVLKPKFPVRIDTSEIKDMVEEHLPLITQOOEVLDKETGFLAEAPD 97
DB 14 KKKNDEN-----KKDEERNKTOEHLKEIM-KHIVKIEVKGEEAVKRAEKLLEKYPVS 68
QY 98 NYKTMRLSKGFSSKVSLETKDGAYTVHIT-----PGPRKTKIANVGVAIGDILSDGNLAE 153
DB 69 DYLEMYKAIG--GKIIV--DGDTHKHSLEALSSEDKKKIKD---IYG---KDALRE 116
QY 154 YNRNLENKQOPVSGDSFDQDSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLN 213
DB 117 HYYVAKGEYPLVVIQSSDYVENTKA-----LN 146
QY 214 VVDSGRPIAGDPFETGTQRYPQIVSGLARFQPGTPYD--LDLLLDQQALEONGHYS 271
DB 147 VYIEIGKILS-----RDILSKI-----NOPYOKFLDVLNTIKNASDSGQ-- 186
QY 272 GASVQADFRLQGRVPVKVSVTEVKRKHLETKGIRLDSYGLGKIAYDYNNLKNKYIG 331
DB 187 -----DLL-----FTNQLKEHPTDFSVFLEQN-----SNEVQEVFAKAF-- 221
QY 332 SVVWDMKYEYTTLAGISQPRNYGNWTNSVNSRSTTONLEK-----RAFS 379
DB 222 -AYIEPQHRDVLQYAPAFNYMDKFNQEQEINLTRAANADVVSILTCPVAAGECAGPADS 280
QY 380 GGIWVVRAGIDARL-----GAEFLEAGRKIPGSDIDLGNSHATMLTASW--KRLLN 431
DB 281 G-----DALLERNYPTGAFLDGGDV-----SFSTRGTQNTVRELQA 320
QY 432 NVLHPENGIYLDGKTGTTGLFSLSTALIRTSARAGYFFTPENKKLGT-----FIRGQ-- 485
DB 321 HRQLEERYGVVGYHGTGLE---AAQSVIFGGVRA-----RSQDLDAIWRGYIAGDPA 371
QY 486 AGYTVARNADVPSCGLMPRSG-----ASSVRGYELDSIGLAPNGSVLPERALLVGS 538
DB 372 LAYGAQOQEPDARG-RIRNCALLRVVPRSSLPGYRTSLTLAAPEAAGEVER--LIG- 427
QY 539 LEYQLP 544
DB 428 --HPLP 431

RESULT 24

US-09-346-408-11
; Sequence 11, Application US/09346408B
; Patent No. 6338966
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Anderson, Shawn
; APPLICANT: Falco, Carl
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
; FILE REFERENCE: BB-1167-A
; CURRENT APPLICATION NUMBER: US/09/346,408B
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 60/092,833
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-346-408-11

Query Match 3.5%; Score 111; DB 4; Length 463;
Best Local Similarity 22.0%; Pred. No. 0.098;
Matches 108; Conservative 68; Mismatches 202; Indels 114; Gaps 26;

QY 96 PNVTMLRSK---GYFSSKV--SLTEKDGAYTVHITPG-----PRTKIAN 136
DB 26 PONLPLSWRSRYRAGPAARTRCGLIEPDGKLVLEIPEPQRLKRRQALSLPQIKLSK 85
QY 137 VGVAILGILSDGLNLAERYNALENWQOPV-----GSDFDQDSWENSKTSVLGAVTRKGY 191
DB 86 IDIQWV-HVLSSEG-----WASPLKGFMRSEFLOTLPNSLRLDGDSVNNMSV 132
QY 192 PLA-KLGNTRAVNPDTATVDLNVVDSGRPIA-FGDPEITGTQRYPQIVSGLARFQPG 249
DB 133 PIVLAIDDSNNKNIIDSSVAL--VDDKDNPIALINDIYKHNK-EERTARTWGTAPG 189
QY 250 TPV-----DLDDL--LDFQALQNGHYSGASVQADFRLQGRV----- 287
DB 190 LPAEQATHAGNLLIGDLEVIPKHYHGLDR-FRLSPAELRDEETRNADAVFAQL 248
QY 288 --PV-----KVSVEVKRKHLETKGIR--LDSEYGLGKIAYDYNNLKNKYIGSVVWDMK 339
DB 249 RNPVINGHALLMTDTRRLLEMGYKNPVLHLLPGG-----YTKADDVPLEMRMKQ 299
QY 340 YETTLAAGISQPRNYGNWTNSVNSRSTTONLEKRAFSGGIWVVRAGIDARLGAEF 399
DB 300 HEWLEDGVLPETTVVSIFPSPMHYAGTEVO-----WHAKAR--INA--GANF 345
QY 400 LAEGRKIPGSDIDLGNSHATMLTASWKROLLNNVLPENGIYLDGKTGTTGLFSLSTAL 459
DB 346 YTVGRDPAGMGHPL--EKRDYDADHGKVLKSMAPGLERLNLIPFKV-----AA 392
QY 460 IRTSARAGYFFTPENKKLGTFIIRGOAGYTVARNADVPSCGLMFRSGGASSVRGYELDSI 519
DB 393 YDKTQNGMAFFDPSRPQDFL-ISTKMRALAKNKENPPDGFMCPGKGVLYDYI--DSL 449
QY 520 GIAGPNGSVLPE 531
DB 450 -TPSENGRV-PE 459

RESULT 25
US-08-164-292B-26
; Sequence 26, Application US/08164292B
; Patent No. 5820868
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVEC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE

; TITLE OF INVENTION: ADENOVIRUS EXPRESSION VECTOR SYSTEM
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 345 California Street
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94104-2675
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/164,292B
; FILING DATE: 09-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 29310-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 677-7000
; TELEFAX: (415) 677-7522
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 983 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-164-292B-26

Query Match 3.5%; Score 111; DB 2; Length 983;

Best Local Similarity 22.2%; Pred. No. 0.35; Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

QY 3 IKPTALL-----LPALFFPHAYAPADISENKAAGFALFKNSPDTSYKLPKF-----53
DB 73 INPAGLLETNDLASAVFPPLASDEAGNVTLMNSDGLYTKDKL-----AVKVGPLSLDSN 128
QY 54 -PVRIDTOD-----SEIKDMVEEHLPLTQQOEEVLDKEQTFGLAEAPDNVKTMLRSKY 108
DB 129 NALQVHTGDLTVTDKVKSLNTQAPLSTTSAGLSLLGLPSHLGEEERLTVM-----181
QY 109 FSSKVSLETKDGAYTVHITPGPRTKIANVGVAIGDILSDGNLAAYRNALENWQOPVGS 168
DB 182 ---GAGLQTSNNALAVKVGSGITVDAQNQLAASLGDL-----216
QY 169 DFDQDSWENSKTSVLG---AVTRKGYPLAKIGNTRAANPDATVDLNVVDSGRPIAFG 225
DB 217 ----ESRDNKTVVAKAGPLTITNQLTVA-TGN-GLOVNP-EQLQLNITACGLNFANN 269
QY 226 DFEITGTORYPQIYVSGIARFQPGT-----PYD-LDL-----LLDFOQAL 264
DB 270 SLAV-----ELSGGL-HFPFGQNOVSLYPGCGIDIRDNRTVTPAGPGLRMLNHOLAV 320
QY 265 EQNGHYSGASVQADRLQGDGRVPVKYSVT---EVKHKLETGIRLDS-----EVLGG 315
DB 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGLTGDDSGRSVVRTGRL 372
QY 316 KIAYDYNNLFN-KG-YIGSVVMDKYEYTTLAAGISOPRNYGNWYTSNVSYNRSTQNL 373
DB 373 RVANGOVQIFSGRGTAGT-----DSSLTNIRAPLQFSGPALTASLQSGSPITYNS 424
QY 374 EKRAF-----SGGIWVVRDRAGIDARLGAER-----LAEGRKIPGSDIDLQ-----NS 416
DB 425 NNGTGLSLGPGWVNDQNRLOVNCAGLVFGQGNLNVNPLADPLAISLSLGLPGLTQ 484
QY 417 HATMLTASWKROLLNNVHPENGHYLDCKIGTTLGTFLSALTIRTSARAGYFFTPENKK 476
DB 485 ASNAULTLS-----LNGLEFSN-QAVALKAGRGL-RPESSQALESSTLVGNGLT-----532

QY 477 LGFTFIIRGOAGYTV-ARDN---ADVPGLMFRSGGASSVRGVELDSIGLAGPNGSLVP---530
DB 533 L7DVTVIRPNLGDGLEVRDNKIIVKLGANLRFENGAVTA-----GTNPSAPEAPT 583
QY 531 ---ERALLVGSLEYQLPFRTRLTSLGAVFHD-----MGDA-AANFKRMKLRKHSGLGVR 578
DB 584 LTAEPPLRASNSHLQLSL---EGLVYVNNALALQLDGMVNVQHGLTLRVSGSLOMR 638

RESULT 26
US-08-845-623-26
; Sequence 26, Application US/08845623A
; Patent No. 6001591
; GENERAL INFORMATION:
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: BOVINE ADENOVIRUS 3 GENOME
; FILE REFERENCE: 29310202120
; CURRENT APPLICATION NUMBER: US/08/845,623A
; CURRENT FILING DATE: 1997-04-25
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
; US-08-845-623-26

Query Match 3.5%; Score 111; DB 3; Length 983;

Best Local Similarity 22.2%; Pred. No. 0.35; Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

QY 3 IKPTALL-----LPALFFPHAYAPADISENKAAGFALFKNSPDTSYKLPKF-----53
DB 73 INPAGLLETNDLASAVFPPLASDEAGNVTLMNSDGLYTKDKL-----AVKVGPLSLDSN 128
QY 54 -PVRIDTOD-----SEIKDMVEEHLPLTQQOEEVLDKEQTFGLAEAPDNVKTMLRSKY 108
DB 129 NALQVHTGDLTVTDKVKSLNTQAPLSTTSAGLSLLGLPSHLGEEERLTVM-----181
QY 109 FSSKVSLETKDGAYTVHITPGPRTKIANVGVAIGDILSDGNLAAYRNALENWQOPVGS 168
DB 182 ---GAGLQTSNNALAVKVGSGITVDAQNQLAASLGDL-----216
QY 169 DFDQDSWENSKTSVLG---AVTRKGYPLAKIGNTRAANPDATVDLNVVDSGRPIAFG 225
DB 217 ----ESRDNKTVVAKAGPLTITNQLTVA-TGN-GLOVNP-EQLQLNITACGLNFANN 269
QY 226 DFEITGTORYPQIYVSGIARFQPGT-----PYD-LDL-----LLDFOQAL 264
DB 270 SLAV-----ELSGGL-HFPFGQNOVSLYPGCGIDIRDNRTVTPAGPGLRMLNHOLAV 320
QY 265 EQNGHYSGASVQADRLQGDGRVPVKYSVT---EVKHKLETGIRLDS-----EVLGG 315
DB 321 A-----SGDGLVHSDTL---RLKLSHGLTFENGAVRAKLGPGLTGDDSGRSVVRTGRL 372
QY 316 KIAYDYNNLFN-KG-YIGSVVMDKYEYTTLAAGISOPRNYGNWYTSNVSYNRSTQNL 373
DB 373 RVANGOVQIFSGRGTAGT-----DSSLTNIRAPLQFSGPALTASLQSGSPITYNS 424
QY 374 EKRAF-----SGGIWVVRDRAGIDARLGAER-----LAEGRKIPGSDIDLQ-----NS 416
DB 425 NNGTGLSLGPGWVNDQNRLOVNCAGLVFGQGNLNVNPLADPLAISLSLGLPGLTQ 484
QY 417 HATMLTASWKROLLNNVHPENGHYLDCKIGTTLGTFLSALTIRTSARAGYFFTPENKK 476
DB 485 ASNAULTLS-----LNGLEFSN-QAVALKAGRGL-RPESSQALESSTLVGNGLT-----532

QY 477 LGFTFIIRGOAGYTV-ARDN---ADVPGLMFRSGGASSVRGVELDSIGLAGPNGSLVP---530

Db 533 LTDTVIRPNLGDGLEVDRNKLIVKLGANLRFENGAVTA-----GTVPNSAPEAPPT 583
QY 531 ---ERALLVGSLEYQLPFTTSLGAVPHD-----MGDA-AANFKRMKLGKHSGLGVR 578
Db 584 LTAEPPLRASNSHLQLSLS---ECLVYHNNALALQDGMGMEVNOHGLTLRVGSGLOMR 638

RESULT 27
US-08-815-927-26
; Sequence 26, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVIC, LUDVIK
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
; TITLE OF INVENTION: VECTOR SYSTEM
; FILE REFERENCE: 293102002101
; CURRENT APPLICATION NUMBER: US/08/815,927
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
US-08-815-927-26

Query Match 3.5%; Score 111; DB 3; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

QY 3 IKPTALL-----LPALFFPHAYAPADLSNKAAGFALFKNKSPTDESVKLKPKF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL-----AVKVGPGLSLDSN 128
QY 54 -PVRIDTOD---SEIKDMVEEHLPLITQOOEVLDEKQOTGLAEAPDNVKTMLRSKY 108
Db 129 NALQVHTGCGLTVTDDKVSUNTOAPLSTTSAGLSLLGSPSLHGLEERLTVNT----- 181
QY 109 FSSKVSLEKDGAVTVHITPGRTKIANGVAILGDLSDGSLAEYRNALENWQOPVGS 168
Db 182 ---GAGLQISNNALAVKVGSGITVDAQNQLAASLGDL----- 216
QY 169 DFDQDSWNSKTSVLG---AVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFG 225
Db 217 ---ESRDNKTVVWAGPGLTITNOALTVA-TGN-GLOVAPE-GQLQLNITAGQGLNFANN 269
QY 226 DFEITGTORYPEQIVSGLARFPQGT-----PYD-LDL-----LLDFOQAL 264
Db 270 SLAV-----ELGSGL-HFPQCQNOVSLYPGDGIDIRDRNRTVTPAGPGLRMLNQLAV 320
QY 265 EQNGHYSGASVQADFDRLQDGRVPVKSVT---EVKRHKLETGIRLDS-----EYGLGG 315
Db 321 A-----SDGLEVHSIDL---RLKLSHGLTFENGAVRAKLGFGLTDDSGRSVVTGRGL 372
QY 316 KIADYYNLFN-KG-YIGSVVMDKYETTTLAAGISQPNRYRGNWYTSNVSNRSTTONL 373
Db 373 RVANGQVQIFSGRGTACT-----DSSLTLNIRAPLQFSGPALTASLOGSGPITYNS 424
QY 374 EKRAF---SGGIWVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTGLSIGPGMWVDQNRLOVNPAGLVPOGNNLVNPLADPLAIDSLSLGLPGLTQ 484
QY 417 HATMLTASWKKROLLNNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKK 476
Db 485 ASNALTL-----LNGLEFSN-QAVALKAGRGL-RFESSQALESSLTVGNGLT----- 532
QY 477 LGTFIIRGAGYTV-ARDN---ADVPSGLMFRSGGASSVRGYELDSIGLAGNGSVLP-- 530

Db 533 LTDTVIRPNLGDGLEVDRNKLIVKLGANLRFENGAVTA-----GTVPNSAPEAPPT 583
QY 531 ---ERALLVGSLEYQLPFTTSLGAVPHD-----MGDA-AANFKRMKLGKHSGLGVR 578
Db 584 LTAEPPLRASNSHLQLSLS---ECLVYHNNALALQDGMGMEVNOHGLTLRVGSGLOMR 638

RESULT 28
US-09-103-330-26
; Sequence 26, Application US/09103330A
; Patent No. 6319716
; GENERAL INFORMATION:
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: ISOLATION OF MUTANTS IN THE E3 REGION OF THE
; TITLE OF INVENTION: BOVINE ADENOVIRUS GENOME AND THEIR USE IN VACCINES
; FILE REFERENCE: 293102002121
; CURRENT APPLICATION NUMBER: US/09/103,330A
; EARLIER FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 08/880,234
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: 08/164,292
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
US-09-103-330-26

Query Match 3.5%; Score 111; DB 4; Length 983;
Best Local Similarity 22.2%; Pred. No. 0.35;
Matches 146; Conservative 78; Mismatches 260; Indels 174; Gaps 36;

QY 3 IKPTALL-----LPALFFPHAYAPADLSNKAAGFALFKNKSPTDESVKLKPKF----- 53
Db 73 INPAGLLETNDLASAVFPPLASDEAGNVTLNMSDGLYTKDNKL-----AVKVGPGLSLDSN 128
QY 54 -PVRIDTOD---SEIKDMVEEHLPLITQOOEVLDEKQOTGLAEAPDNVKTMLRSKY 108
Db 129 NALQVHTGCGLTVTDDKVSUNTOAPLSTTSAGLSLLGSPSLHGLEERLTVNT----- 181
QY 109 FSSKVSLEKDGAVTVHITPGRTKIANGVAILGDLSDGSLAEYRNALENWQOPVGS 168
Db 182 ---GAGLQISNNALAVKVGSGITVDAQNQLAASLGDL----- 216
QY 169 DFDQDSWNSKTSVLG---AVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFG 225
Db 217 ---ESRDNKTVVWAGPGLTITNOALTVA-TGN-GLOVAPE-GQLQLNITAGQGLNFANN 269
QY 226 DFEITGTORYPEQIVSGLARFPQGT-----PYD-LDL-----LLDFOQAL 264
Db 270 SLAV-----ELGSGL-HFPQCQNOVSLYPGDGIDIRDRNRTVTPAGPGLRMLNQLAV 320
QY 265 EQNGHYSGASVQADFDRLQDGRVPVKSVT---EVKRHKLETGIRLDS-----EYGLGG 315
Db 321 A-----SDGLEVHSIDL---RLKLSHGLTFENGAVRAKLGFGLTDDSGRSVVTGRGL 372
QY 316 KIADYYNLFN-KG-YIGSVVMDKYETTTLAAGISQPNRYRGNWYTSNVSNRSTTONL 373
Db 373 RVANGQVQIFSGRGTACT-----DSSLTLNIRAPLQFSGPALTASLOGSGPITYNS 424
QY 374 EKRAF---SGGIWVRDRAGIDARLGAEP-----LAEGRKIPGSDIDLG-----NS 416
Db 425 NNGTGLSIGPGMWVDQNRLOVNPAGLVPOGNNLVNPLADPLAIDSLSLGLPGLTQ 484
QY 417 HATMLTASWKKROLLNNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKK 476
Db 485 ASNALTL-----LNGLEFSN-QAVALKAGRGL-RFESSQALESSLTVGNGLT----- 532


```
QY 532 RALLVCS--LEVQLPFT-----RTLGAVFHDMG---DAAANFKRMKLGKHS 574
Db 860 --TLGSDFDHQPRLIKKIDVINSNLTAGNIYNIAGNLTVESNANFK----- 908
QY 575 LGRWFSPLEPESFDIAYGHSDKKIRWHISL---GTRF 609
Db 909 -----AITNFTNVG--GLFDNKGNSNISIAKGARF 938

RESULT 31
US-09-376-330-18
; Sequence 18, Application US/09376330
; Patent No. 6399321
; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Beigeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGGT
; FILE REFERENCE: 2139-9*US*
; CURRENT APPLICATION NUMBER: US/09/376,330
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: S. cerevisiae KRE5
US-09-376-330-18

Query Match 3.4%; Score 109; DB 4; Length 1365;
Best Local Similarity 20.7%; Pred. No. 0.92;
Matches 124; Conservative 93; Mismatches 212; Indels 170; Gaps 33;

QY 41 SPDTESVKKPKFPVRIDTQD--SEIKDMVEEHLPLITQOOEVLDEKQTFGLAEAPONV 99
Db 208 SLDGKSVPLTHPLTQNGSRMSSIPQLKKILYVPEILVGADNDQLHLEPEEL 267
QY 100 KTWLSKGYFSSKVSLEKDGAYVHTPG-----PRTKIANGVAILGDIILS-- 147
Db 268 REIDLRTVLSIEFVQYKKOTATLNTFKSIVNPNFLISKQLIKVSSVN---KDIITSN 323
QY 148 -----DGNLAEYRNALENWQPVGSDFDQDSWENSKTSVLGAVTRKGYPLAKLGT 199
Db 324 EELNSKGFDMGLGLYING-QNWKITSPTY-----NLLTALKTEYOSLLKITNL 372
QY 200 RAAVNPDTATVDLNVVDSGRPIAGDFEITGTQRPQEVISGLARFQPTPYDLDDLID 259
Db 373 LQELPESKILDSKFLNKK-----FSQFSL-----GKLQNLQP-IRMDLHTIPG 415
QY 260 FOQA-----LEONSHYSGA--SVQADFRLQDRVPVKVSVTEVKRH-----K 300
Db 416 FSESVIYENDIESDPQYDELVNSVQAFDKSKEGELP-----EIKQNMSELIIFVIDPAR 469
QY 301 LETGIRLDSEY--GLGGKIADYNNLNKGY---IGSVVMDMKYETTLAAGISQPNRYR 355
Db 470 LE-----DSEVKEALGGLVR--AVNVVSGQTPQVRGLLPFSSDSKSVY----- 511
QY 356 GNYTWSNYSYN-RSTQNL-EKRAFSGGIWYVRDRAGIDARLGAELAEGRK-----IPGS 409
Db 512 -----NKIYELKNSTDNLTLSKF-----LETMLLADGLSANAKHSHIPVP 553
QY 410 DI-----DLGNSHATMTASWKROLLANVLHP--ENCHYLDGKIGITGLTFLSSALTRT 462
Db 554 DVPHLDELQIDETS-----IINGEYIPFRKNWNYLIQAV-----IKKDTFTFRK 599
QY 463 SARAGYFTTPENKKLGTFIRGOAGYVVA--RONADVPGLMFRSGGASVRYGELDSTG 520
Db 600 ELSNS---SPKNKQIS---VRDLLHYKSANLRHKNKYPN--YFADSVYSVNNTALE--- 648
```

```
QY 521 LAGPNSVLPER-ALLVCSLEYQLPFTRTLTSAGVAFHDMGDAANFKRMKLGKHSGLVGR 578
Db 649 -----SVCSEIRGYITKNEEYNLLHTITL-----VDDFGSIHALKRLRLNLLHTSPVGVGR 697

RESULT 32
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 3.4%; Score 108.5; DB 4; Length 1375;
Best Local Similarity 19.7%; Pred. No. 1;
Matches 134; Conservative 87; Mismatches 269; Indels 189; Gaps 34;

QY 30 KAAGTALFKNKSPTDESVKLPKFPVRIDTQDSEIKDMVEEH-----LPLITQOOEVE-- 82
Db 729 RTSGVAVIEGNNPS-----LRLKASDRVVMVGMGAHKNQAYRPLLLTTDNGIKA 777
QY 83 --LDKEQTGFL--AEEAPDNVKTMLRSKGYFSSKVS-----LTKDGAYTVHIT 127
Db 778 YHSDQEAAGLVRYTNDRGELIFTADIKGYANPQVSGYLVGVVPGAAADQDVVAASTA 837
QY 128 PGPRTKIANGVAILGDIILSDG-----NLAEYRNA-----LENWQOPVGSDFDQDS 174
Db 838 PSTDKSKVHQAALDSRVMEFGSNFOAFATKKEEYTNVVIKNNVDKFAEWGVTDFEMAP 897
QY 175 WENSKT--SVLGAVTRKGYPLAKLGNTRAADVNPDTATVDLNVVDSGRPIAGF----- 225
Db 898 QYVSSDTGDSFLSDVIONGYAF-----TDYDLGI---SKPNKYGTADDLVK 940
QY 226 -----DPEITGTQRYPE-QIVSGLARFQPTP-----YDLDDLDF 260
Db 941 ATKALHSKGIKVMADWVPDOMYALPEKEVVTATRVDRKYGTFPVAGSOKNTLYVVD---GK 997
QY 261 QOALEQNGHYSGA-----SVQADFRLQDRVPVKVSVTEVKRHKLETGIRLDSEYLGKGI 317
Db 998 SSGKDOAKYKAGFLEELQAKYP-----FARKQISTGVPMPSVKIKQWS 1044
QY 318 A--YDYNLNFKNKYGTSVVDWM--DKYETTLAAGISQPNRN-YKGNWTNSN----- 362
Db 1045 AKYFNGTNLGRG--AGYVLKQDQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVFDGK 1103
QY 363 --VSVNRSTQNLKRAF--SGGIWYVRDR-----AGIDARLGA--FLAEGRKIPGSI 411
Db 1104 GYVYV--STSGQAKNAFISLGNWNYFDNNGYVWTCAQISNGANYFSLNGIOLRNAY 1161
```



```
QY 392 DARLGAFLAERKIPGSDIDLGNHSHATMLTASWKROLLNNVLHPENGHYLDKIGTTIG 451
Db 287 KYLDA---AHGAWL-----NOYADQ-TANVKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKTFTIIRG-QAGYTVARNADVPGLMFRSGGASS 510
Db 326 TNVSNQYSIESEYK---YHQNLRALSKGVRLKFTVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLGORPKGNPNPGSMP 401

RESULT 35
US-09-286-691-2
; Sequence 2, Application US/09286691
; Patent No. 6190189
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96
; CURRENT APPLICATION NUMBER: US/09/286,691
; PRIOR FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,883
; EARLIER FILING DATE: 1996-10-04
; EARLIER APPLICATION NUMBER: PCT US97/18008
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-286-691-2

Query Match 3.4%; Score 108; DB 4; Length 459;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 90; Conservative 63; Mismatches 193; Indels 94; Gaps 19;

QY 141 ILGDILSDGNLAEEYRNALNNKQOPVGS-----FQDQSWNSK 179
Db 6 VLATFATGALASEC-----HWQPCCKDCTVYTTDEGKWGVNNDMCMIDNRRCSSNN 60
QY 180 TSVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQI 239
Db 61 NNCSSSITSGQYPCCSNNCKVEYTDN---DGKWGVNENNWCGISNCGGQOQOQTOP 116
QY 240 VSGLARFOPGTPYDLDLLDFQQALEONGHYSGASVQADFDRLQGD-----RVPVK 290
Db 117 TQPTQPTQPTQSSDNF---FENEIYSNYKFGQ-EVDISIKKLNGDLKAKAEKVVPTA 172
QY 291 V-----SVTEVKRHKLETGR-----LDSEYGLGKTIAYDYNLFNKGYIG 331
Db 173 VMLWDGAPQEPYRLOAGNKTVFVLYMIPTRDCGANASAGGSATIDKY-----KGYN 228
QY 332 SYVWMDKYETTTLAAGISOPRNYRGNTWTSNVSNRSTTONLEKRAFSGGIWTVRDRAGI 391
Db 229 NIYNTSNQYKNSKIYVILEPDTI-GNLVTNN-NDNCNRVNRNMHKQALSAISKFGTQSHV 286
QY 392 DARLGAFLAERKIPGSDIDLGNHSHATMLTASWKROLLNNVLHPENGHYLDKIGTTIG 451
Db 287 KYLDA---AHGAWL-----NOYADQ-TANVKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKTFTIIRG-QAGYTVARNADVPGLMFRSGGASS 510
Db 326 TNVSNQYSIESEYK---YHQNLRALSKGVRLKFTVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLGORPKGNPNPGSMP 401

RESULT 37
US-08-613-009A-10
; Sequence 10, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
```

```
RESULT 36
US-09-687-147-2
; Sequence 2, Application US/09687147
; Patent No. 6268198
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Cellulases and Coding Sequences
; FILE REFERENCE: 42-96a
; CURRENT APPLICATION NUMBER: US/09/687,147
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/027,883
; PRIOR FILING DATE: 1996-10-04
; PRIOR APPLICATION NUMBER: PCT US97/18008
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 09/286,691
; PRIOR FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Orpinomyces sp. PC-2
US-09-687-147-2

Query Match 3.4%; Score 108; DB 4; Length 459;
Best Local Similarity 20.5%; Pred. No. 0.18;
Matches 90; Conservative 63; Mismatches 193; Indels 94; Gaps 19;

QY 141 ILGDILSDGNLAEEYRNALNNKQOPVGS-----FQDQSWNSK 179
Db 6 VLATFATGALASEC-----HWQPCCKDCTVYTTDEGKWGVNNDMCMIDNRRCSSNN 60
QY 180 TSVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTQRYPEQI 239
Db 61 NNCSSSITSGQYPCCSNNCKVEYTDN---DGKWGVNENNWCGISNCGGQOQOQTOP 116
QY 240 VSGLARFOPGTPYDLDLLDFQQALEONGHYSGASVQADFDRLQGD-----RVPVK 290
Db 117 TQPTQPTQPTQSSDNF---FENEIYSNYKFGQ-EVDISIKKLNGDLKAKAEKVVPTA 172
QY 291 V-----SVTEVKRHKLETGR-----LDSEYGLGKTIAYDYNLFNKGYIG 331
Db 173 VMLWDGAPQEPYRLOAGNKTVFVLYMIPTRDCGANASAGGSATIDKY-----KGYN 228
QY 332 SYVWMDKYETTTLAAGISOPRNYRGNTWTSNVSNRSTTONLEKRAFSGGIWTVRDRAGI 391
Db 229 NIYNTSNQYKNSKIYVILEPDTI-GNLVTNN-NDNCNRVNRNMHKQALSAISKFGTQSHV 286
QY 392 DARLGAFLAERKIPGSDIDLGNHSHATMLTASWKROLLNNVLHPENGHYLDKIGTTIG 451
Db 287 KYLDA---AHGAWL-----NOYADQ-TANVKEILNNA-----GSGKL-RGIS 325
QY 452 TFLSSTALIRTSARAGYFFTPENKKTFTIIRG-QAGYTVARNADVPGLMFRSGGASS 510
Db 326 TNVSNQYSIESEYK---YHQNLRALSKGVRLKFTVDTSRNGANV-EGAFNASGTWCN 381
QY 511 VRGYELDSIGLAGPNSVLP 530
Db 382 FKAGLGORPKGNPNPGSMP 401

RESULT 37
US-08-613-009A-10
; Sequence 10, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
```

```

: APPLICANT: Du, Run-Pan
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H
: TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/613,009A
: FILING DATE: 08-MAR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24973
: REFERENCE/DOCKET NUMBER: 1038-542
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 682 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-613-009A-10

```

```
Query Match      3.4%; Score 107.5; DB 3; Length 682;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 112; Conservative 69; Mismatches 202; Indels 171; Gaps 26;
```

QY	33	GFALEKNNKSPDFTSEVKLAKPKFPVRIDTQDSIEIKOWEEHPLPLITQQOEEVLOKEQTGFPLA	92
		: : : : : : : : : :	
Db	83	GMALESINLHNQDPTLDKEKNIITLDGK-KOVAEGKKSPLPFSLDVENKLLD-----GYTA	137
		: : : : : : : : : :	
QY	93	EAPADNWKMLRSKGYPSSKYSVLFEKDGAVTVHTTPGPRTKIANVGVAILOILSDGNLA	152
		: : : : : : : : : : : : :	
Db	138	-----KMNVADKNAL-----GDRIKKNKEI-----SDEELA	164
		: : : : : : : : : :	
QY	153	EYYRNALENWQPVGSDFDQ---DAWEN----SKTSVLCAVTRK-----GYPLAKLGNTPR	200
		: : : : : : : : : : : :	
Db	165	KQIKEAVRK-----SHEQQVLSLENKIFHSNDGGTKATTROLKYVDVGYGLANDGN--	217
		: : : : : : : : : :	
QY	201	AANVPOTATVLNVVVDDGRPIAFGEDEPTEIGTRYPEQIVSGLARFGQPPYDLDLLDLDF	260
		: : : : : : : : : : :	
Db	218	-----YLHVTKDKLNULG-PV--GGVFYNQTTTAAE-----LPT	248
		: : : : : : : : : : :	
QY	261	OQALEONGHYSGASVOADFDRLOGDRPVKVSYTEV-KRHKLETGIRLSDXSEGLCGKIAY	319
		: : : : : : : : : : :	
Db	249	QDAVKYKGHW-----DF-----MIDVANRRNRSEVKENSQAGW-----	282
		: : : : : : : : : : :	
QY	320	DYYNLNKGYIGSVVMDMKYETTLAAGISOPRYNGNYWTWSVYNRSTTONLEKRAFS	379
		: : : : : : : : : : :	
Db	283	-YYGASSK-----DEYNRLLTREDSAPDHSGEY-----GHSSEFTVNFREKILT	326
		: : : : : : : : : : :	
QY	380	GGIWTYVRADRIDARLGAELAEGRKIPGSDIDL---GNSHATMLTASAKRWOLLNVLHP	436
		: : : : : : : : : : :	
Db	327	GKLF-----SNLQRHRKGNVTERTYDIDANIHGFRFGSATASNKDTSRKPPTS	377
		: : : : : : : : : : :	
QY	437	ENGHYLDGXIGITTLGTFLSSLTALTITSARAGYPTPENKKLGLFIIRGOAGYTVARDNA-	495
		: : : : : : : : : : :	
Db	378	DANNRLEGGYGPKGEEL-----ACKFTINDONKLGVYF---GAKRESKAPEKTE	423
		: : : : : : : : : : :	
QY	496	---DVPSGMLFRSGGASSVRGY---ELDSIGLAGPGSVLPERALLVGSLEYOLPTRL	549

```

Db      424  AILDVALCTFNTSNATFTFTTEKQLDNFGNA-----KKLVLGSTVIDLVPTDAT 474
QY      550  SGAVFHDMDGAARAN 563
Db      475  KNEFTKDKPESATN 488

RESULT 38
US-08-778-570B-12
: Sequence 12, Application US/08778570B
: Patent No. 6437096
: GENERAL INFORMATION:
: APPLICANT: Myers, Lisa E
: APPLICANT: Schryvers, Anthony B
: APPLICANT: Harkness, Robin E
: APPLICANT: Loomore, Sheena M.
: APPLICANT: Du, Run-Pan
: APPLICANT: Yang, Yan-Ping
: APPLICANT: Klein, Michel H
: TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1K7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/778,570B
: FILING DATE: 03-JAN-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24973
: REFERENCE/DOCKET NUMBER: 1038-664
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 682 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-778-570B-12

```

Query Match	3.4%	Score 107.5	DB 4	Length 682
Best Local Similarity	20.2%	Pred. No. 0.4		
Matches 112	Conservative 69	Mismatches 202	Indels 171	Gaps 26
Qy	33	GFALFNKSPDTESVLKKPFVRIOTDSEIKMVEEHLPLITQOEEVLDKEQTGFLA	92	
		: : : : :		
Db	83	GMALSKLIHNFODPTLDENKIITLDGR-KQVAGRKSPFLSDUVENKLLD----	137	
		: : : : :		
Qy	93	EEAPDNVKTMLRSKGYFSSKSLTEBKDAYTVHITPGPRTKIANVGVAIIGDILSDGUA	152	
		: : : : :		
Db	138	-----KMMVADKNAI-----GDRIKKNKEI-----SDEELA	164	
Qy	153	EYRNALENNQQPVGSDFDQ-----DSWEN-----SKTSVLGAVTRK-----GYPLAKLGNT	200	
		: : : : :		
Db	165	KQIEAVRK-----SHEFOOVLSSLEKNKIFISNDCITKATYTRDLKYVDYGYGLANDGN--	217	
Qy	201	AAVNPOTATVDLNVVVDGSRPLAFGDFEITQYRPEQIVSGLARFQPGTPPYDLDLLD	260	
		: : : : :		
Db	218	-----VLTVKTDKLNWILQ-PV--GGVYNGTITTAKE-----LPT	248	

```

QY 261 QOALSEQNGHYSGASQVADFRLQGRVPVKKVSTEV-KRHKLETGIRDSEYGLGKGIAY 319
Db 249 QDAVYKCGHW-----DF-----MTDVA NRNRNFSEVKENSOAGW-----282
QY 320 DYNNLFNKGYIGSVWMDKYEYTTLAAGISOPRNTYRGNYWTSNVSYNRSTTQNLKRAFS 379
Db 283 -YYGASSK-----DEYNRLTKEDSAPDGHSGEY-----CHSSEFTVNFKEKKLT 326
QY 380 GGIWYVRDRAGIDARLCAEFLAERKIPQSDIDL---GNSHATMLTATSWKROLLNNVLHP 436
Db 327 GKLF-----SNLDRHKGNWTKTERYDIDAHIGNRFRGSATASNKNDTSKHPFTS 377
QY 437 ENGHYLDGKIGTTLGTLSLSTALIRTSABAGYFFFTPENKKGITFIIRGOAGYTVARDNA- 495
Db 378 DANNRLEGFCGPKAGEEL-----AGKFLTDNKNLFCVF---CAKRESKAEREKE 423
QY 496 ---DVPSCGLMFRSGGASSVRGY---ELDSIGLAGPNSGVLPERALLVGSLEYQLPFTRTL 549
Db 424 AILDAYALGTFTSNATITFTPTFEKOLDNFGNA-----KKLVLGSTVIDLVPTDAT 474
QY 550 SGAVFHDMDGDAAN 563
Db 475 KNEFTKDKPESATN 488

RESULT 39
US-09-059-584-12
: Sequence 12, Application US/09059584
: Patent No. 6440701
: GENERAL INFORMATION:
: APPLICANT: Myers, Lisa E
: APPLICANT: Schryvers, Anthony B
: APPLICANT: Harkness, Robin E
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Du, Run-Pan
: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H
: TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/059,584
: FILING DATE: 14-APR-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/778,570
: FILING DATE: 03-JAN-1997
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24973
: REFERENCE/DOCKET NUMBER: 1038-794
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 682 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-059-584-12

```

```

: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 702 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-867-941-25

Query Match      3 48: Score 107.5; Db 2; Length 702;
Best Local Similarity 20.2%; Pred. No. 0.42;
Matches 112; Conservative 69; Mismatches 202; Indels 171; Gaps 26;

QY 33 GPALPKNSPDTSYKLPKPPVRIDTQDSEIKDMVEEHLPLITQQQEEVLDKEQTGFIA 92
DB 403 GMAISKINLHNRQDTPLDEKNIITLDGK-KQVAEGKKSPLPESLQVENKLLD---GVIA 157
QY 93 BEAPDNVKTMLRSKGYSSKVSLETKDGAVTVHITPGPKTIANYGVAILGDLISDGNLA 152
DB 158 -----KMNADKNAI-----GDRIKKGNKEI-----SDEELA 184
QY 153 EYRNALENMOOPVGSDFDQ--DSWEN-----SKTSVLGAVTRK-----GYPLAKLGNT 200
DB 185 KQIKEAVRK-----SHEFQVLSLENKIFHSNDGTTKATTROLKYVDYGYVYLANDGN-- 237
QY 201 AAVNPDTATVLDNVVDSGRPIAFGDFEITGTQRYPEQIVSGLARFQPGTPYDLDLLDF 260
DB 238 -----YLTVTDKLWNLG-PV--GGVYNGTTAKE-----LPT 268
QY 261 QOALEQNGHYSGASVQADDFRLOQDRVPVKVSVTEV-KRKLETGIRLDSEYGLGCKIAY 319
DB 269 QDAVKYKGNW-----DF-----MTDVANRRNRFSEVKENSQAGW----- 302
QY 320 DYINLNFNGYIGSVVWDMDKYETTLAAGISOPRNYRGNYWTSNVSYNRSTTONLEKRAFS 379
DB 303 -YYGASSK-----DEYNRLLTKEDSAPDGHSGEY-----GHSSEFTVNFKEKKLT 346
QY 380 GGIWTVRRAGIDARLAEGFLAEGRKIPGSDIDL---GNSHATMLTASWKROLLNNVLHP 436
DB 347 GKLF-----SNLQDRHKGNVTKTERYDIDANTHGNRFRGSATASNKNDTSKHPPTS 397
QY 437 ENGHVLDGKIGTTLGTELSTALIRTSARAGYEFTPENKKLGTFIIRGOAGYTVARDNA- 495
DB 398 DANNLEGGFYGPKEEL-----AGKELTNDNKLFGVF---GAKRESKAEKTE 443
QY 496 ---DVPSCLMFRSGAGSVRGY---ELDSIGLAGPNSGVLPERALLVGSLEYQLPFTFTL 549
DB 444 AILDAYALGTTNTSNATFTPTTEKQLQNFNA-----KKLVLGSTVIDLVPTDAT 494
QY 550 SGAVFHDMDGAAAN 563
DB 495 KNEFTKDKPESATN 508
```

Search completed: November 9, 2002, 01:21:40
Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2002, 01:17:09 ; Search time 21 seconds
(without alignments)
2787.898 Million cell updates/sec

Title: US-09-857-669-2
Perfect score: 3173
Sequence: 1 MMKPTALLPALFFPHAY.....IAYGHSDKIRWHISLGRF 609

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues 283224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3153	99.4	615	2 AB2025	probable outer mem
2	3136	98.8	635	2 G81003	conserved hypothe
3	449	14.2	579	2 F83327	conserved hypothe
4	379.5	12.0	577	2 S56445	hypothetical 64,8K
5	379.5	12.0	577	2 F91278	hypothetical prote
6	379.5	12.0	577	2 F86119	hypothetical prote
7	374	11.8	578	2 B64012	hypothetical prote
8	368.5	11.6	578	2 AE0428	probable exported
9	355.5	11.2	577	2 AD1055	probable exported
10	348	11.0	582	2 F82064	conserved hypothe
11	323.5	10.2	641	2 H97672	hypothetical prote
12	320	10.1	617	2 AF2897	conserved hypothe
13	306.5	9.7	617	2 G82706	conserved hypothe
14	290.5	9.2	623	2 A13488	outer membrane pro
15	279	8.8	833	2 AF2089	hypothetical prote
16	272.5	8.6	628	2 B87448	conserved hypothe
17	253	8.0	797	2 D82000	outer membrane pro
18	246	7.8	778	2 G81228	outer membrane pro
19	246	7.8	795	2 C70412	outer membrane pro
20	246	7.8	795	2 AC0129	probable surface a
21	227	7.2	803	2 B82099	surface antigen VC
22	226	7.1	803	2 AB0530	outer membrane pro
23	217	6.8	810	2 A6742	hypothetical prote
24	217	6.8	810	2 C90651	hypothetical prote
25	217	6.8	810	2 C85502	hypothetical prote
26	215.5	6.8	797	2 H83190	probable outer mem
27	211	6.6	676	2 AE2417	hypothetical prote
28	210.5	6.6	769	2 F87486	outer membrane pro
29	209.5	6.6	797	2 JC4078	protective surface

30 209.5 6.6 808 2 F64102 protective surface
31 206 6.5 739 2 A81430 outer membrane pro
32 202.5 6.4 861 2 S77409 hypothetical prote
33 199 6.3 792 2 H81693 outer membrane pro
34 197.5 6.2 784 2 E82731 outer membrane ant
35 194.5 6.1 781 2 AH3355 outer membrane pro
36 185 5.8 774 2 D97527 omp1 protein precu
37 185 5.8 774 2 AE2746 group 1 outer memb
38 179 5.6 491 2 AC1816 hypothetical prote
39 174 5.5 906 2 F71910 probable outer mem
40 173.5 5.5 790 2 D86528 omp85 analog [impo
41 173.5 5.5 790 2 D72094 omp85 analog - Chl
42 172.5 5.4 916 2 G64601 protective surface
43 169 5.3 792 2 B71539 probable omp85 ana
44 163 5.1 768 2 D71726 outer membrane pro
45 163 5.1 768 2 B97725 outer membrane pro

ALIGNMENTS

RESULT 1
AB2025
probable outer membrane protein NMA0296 [imported] - Neisseria meningitidis (strain 7
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB2025
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandr
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 2249
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: AB2025
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83602.1; PID:g73
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA0296

Query Match 99.4%; Score 3153; DB 2; Length 615;
Best Local Similarity 99.3%; Pred. No. 8.1e-204;
Matches 605; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MMKPTALLPALFFPHAYAPAAADLSENKAAGFALFKNKPDTESVKLKPFPPVRIQTQ 60
Db 7 MMKPTALLPALFFPHAYAPAAADLSENKAAGFALFKNKPDTESVKLKPFPPVRIQTQ 66
QY 61 DSEIKDMVEEHLPLITQQOBEVLDKQGTGFLAEEAPDNVKTMLRSKGYESSKVSUTEKDG 120
Db 67 DSEIKDMVEEHLPLITQQOBEVLDKQGTGFLAEEAPDNVKTMLRSKGYESSKVSUTEKDG 126
QY 121 AYTTHITGPRTKTANVGVALTGDLSDGNLAEEYRNALENWQOPVGSDFDQDSNENSKT 180
Db 127 AYTTHITGPRTKTANVGVALTGDLSDGNLAEEYRNALENWQOPVGSDFDQDSNENSKT 186
QY 181 SVLGAVTRKGYPLAKLGNRAVNPDVTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 240
Db 187 SVLGAVTRKGYPLAKLGNRAVNPDVTATVDLNVVDSGRPIAFGDFEITGTQRYPEQIV 246
QY 241 SGLARFQGTPTDLDLLDFQOALEQNGHYSASVQADFRLQGDVPPVKVSVTEVKRHK 300
Db 247 SGLARFQGTPTDLDLLDFQOALEQNGHYSASVQADFRLQGDVPPVKVSVTEVKRHK 306
QY 301 LETGTRLOSEYGLGKAIYDYNLNFNGYTGVSVMWMDKYETTLAAGISQPRNRYGNWT 360
Db 307 LETGTRLOSEYGLGKAIYDYNLNFNGYTGVSVMWMDKYETTLAAGISQPRNRYGNWT 366
QY 361 SNVSNRSTTQNLKRAFGSGIWTYVRDAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 420
Db 367 SNVSNRSTTQNLKRAFGSGIWTYVRDAGIDARLGAFLAEGRKIPGSDIDLGNSHATM 426

Qy 421 LTASWKRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
|||||
Db 427 LTASWKRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 486
|||||
Qy 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLE 540
|||||
Db 487 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLE 546
|||||
Qy 541 YQLPFTRTLSSGAVFHDMDGAAANFRMKLKHGSGLVGRWFSPLAPFSFDIAYGHSOKKIR 600
|||||
Db 547 YQLPFTRTLSSGAVFHDMDGAAANFRMKLKHGSGLVGRWFSPLAPFSFDIAYGHSOKKIR 606
|||||
Qy 601 WHISLGTRF 609
|||||
Db 607 WHISLGTRF 615
|||||

RESULT 2
G81003
conserved hypothetical protein NMB2134 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81003
R:Kettellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: G81003
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-635 <TET>
A:Cross-references: GB:AE002561; GB:AE002098; NID:97227384; PIDN:AAF42442.1; PID:9722739
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB2134

Query Match 98.8%; Score 3136; DB 2; Length 635;
Best Local Similarity 98.7%; Pred. No. 1.2e-202;
Matches 601; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MMKPTALLPALFFPPHAYAPADLSENKAAGFALFKNKSPDTESVKLPKPEPVRIIDTQ 60
|||||
Db 27 MMKPTALLPALFFPPHAYAPADLSENKAAGFALFKNKSPDTESVKLPKPEPVRIIDTQ 86
|||||
Qy 61 DSEIKDWEEHLPLITQOQEEVLKDEQGTFLAEAPDNVKTMLRSKGYFSSKVSLETKDG 120
|||||
Db 87 DSEIKDWEEHLPLITQOQEEVLKDEQGTFLAEAPDNVKTMLRSKGYFSSKVSLETKDG 146
|||||
Qy 121 AYVHTIPGPRTKIANVGVAILGDILSDGNLAEYRNALENNMQQPVGSDFDQDQSWNSKT 180
|||||
Db 147 AYVHTIPGPRTKIANVGVAILGDILSDGNLAEYRNALENNMQQPVGSDFDQDQSWNSKT 206
|||||
Qy 181 SVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQIV 240
|||||
Db 207 SVLGAVTRKAYPLAKGNTRQAAVNPDTATADLNVVDSGRPIAFGDFEITGTORYPEQIV 266
|||||
Qy 241 SGLARFQPGTPYDLDLLDFQALBQNGHYSGASVQADFDRLQDGRVPKVSYTEVKRHK 300
|||||
Db 267 SGLARFQPCMPYDLDLLDFQALBQNGHYSGASVQADFDRLQDGRVPKVSYTEVKRHK 326
|||||
Qy 301 LETGIRLDSYGLGGKIAYDYNLKNKYIGSVVMDMDKYETTLAAGISOPRNYGWT 360
|||||
Db 327 LETGIRLDSYGLGGKIAYDYNLKNKYIGSVVMDMDKYETTLAAGISOPRNYGWT 386
|||||
Qy 361 SNYSYNRSTQNLKRAFSGGIWYVRDAGIDARLGAELAEKRIKPGSIDLGNSHATM 420
|||||
Db 387 SNYSYNRSTQNLKRAFSGGIWYVRDAGIDARLGAELAEKRIKPGSAVDLGNSHATM 446
|||||
Qy 421 LTASWKRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 480
|||||

Db 447 LTASWKRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFFTPENKKLGTF 506
|||||
Qy 481 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLE 540
|||||
Db 507 IIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLE 566
|||||
Qy 541 YQLPFTRTLSSGAVFHDMDGAAANFRMKLKHGSGLVGRWFSPLAPFSFDIAYGHSOKKIR 600
|||||
Db 567 YQLPFTRTLSSGAVFHDMDGAAANFRMKLKHGSGLVGRWFSPLAPFSFDIAYGHSOKKIR 626
|||||
Qy 601 WHISLGTRF 609
|||||
Db 627 WHISLGTRF 635
|||||

RESULT 3
F83327
conserved hypothetical protein PA2543 [imported] - Pseudomonas aeruginosa (strain PAO)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83327
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lerbog, K.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83327
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AE004682; GB:AE004091; NID:99948598; PIDN:AAG05931.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA2543

Query Match 14.2%; Score 449; DB 2; Length 579;
Best Local Similarity 22.9%; Pred. No. 2.7e-22;
Matches 144; Conservative 111; Mismatches 28; Indels 84; Gaps 16;

Qy 9 LLPALFFPHAYAPADLSENKAAGFALFKNKSPDTESVKLPKPEPVRIIDTQDSEIKDW 68
|||||
Db 3 LTQGLFRF-----PLAGLLCLSTAFA-----ESOLSVRVTPANAALKANI 43
|||||
Qy 69 EEHLPLITQOQEEVLKDEQGTFLAEAPDNVKTMLRSKGYFSSKVSLETKDG---AYTVH 125
|||||
Db 44 EAVVSGLSGERDEAALQJ-----FRNBAQAEKAAQALGVFOAQIDSEVKGDKPKPLTK 96
|||||
Qy 126 ITPGPRTKIANVGVAILGDILSDGNLAEYRNALENNMQQP-----VGSDFDQDQSWNSK 179
|||||
Db 99 VVPEPVRLLRQVNIQVLGEAAS-----LESFRLPSGKQLKPGAKLNGVVEDAK 147
|||||
Qy 180 TSVLGAVTRKGYPLAKGNTRAAVNPDTATVDLNVVDSGRPIAFGDFEITGTORYPEQI 239
|||||
Db 148 RLQONASRYGFGQGRFSTORLISIDPRAGIADIDLVDSDGQRYTFGKVSDFGDSIIIEEL 207
|||||
Qy 240 VSGLARFQPGTPYDLDLLDFQALBQNGHYSGASVQADFDRLQDQDQDQDQDQDQDQDQ 295
|||||
Db 208 LRRWPFKAGQPYDSELIAELNQLQSSGYFEGVVRDAAPTQAGADGARQAIIPAVRLEA 267
|||||
Qy 296 VKRHKLETGIRLDSYGLGGKIAYDYNLKNKYIGSVVMDMDKYETTLAAGISOPRNYGWT 341
|||||
Db 268 RKDRTMGVLGFSTQDVGARARFNWTRHWNAEGHSLGFESEISAPRONVGA--W-----YE 321
|||||
Qy 342 TTLAAGISOPRNYGWTSNYSYNRSTQNLKRAFSGGIWYVRDAGIDARLGAELAE 401
|||||
Db 322 IPLDPLTDKLRFTSGYQFEDLVDTESKLLTL-----GGEHSHKRPDQGRVSVLNWR 375
|||||
Qy 402 EGRKIPGSDIDLGNSHATMLTASWKRQLLNVLHPENGHYLDGKI-CTTLGTFLSSTAL- 459
|||||
Db 376 EYKLGQD---DSGLSFLMPGIGYSLLETDNKNVDP-SHGIRLQFNVKAKGLAADADVLH 432
|||||
Qy 460 IRTSARAGYFFTPENKKLGTFIIRGOAGYTVARDNADVPGLMFRSGGASSVRGYELDSI 519
|||||

Db 433 VDAKAGLTSEAGGRLG-----BLQGGIATNDYKSLPPSLRFFAGGDSVRYGYDRTL 488
QY 520 GLAGPNSVLPERRALLVGSLEYQLPFTRTLSCGAVFHDMDGDAANFRKMLKHGSLGVRW 579
Db 489 SPENSDGKIGRYWAGSYEQPLAERWRLATFVDQGNAFNSLDLFPFSIKTGVGVRW 548
QY 580 FSLAPFSDIAYG-HSDKKIRHISLG 606
Db 549 VSPVGLRLDLAHLDDGGRFLRSMG 576

RESULT 4

S56445
N:Alternate names: msra-chpbi intergenic region) - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C:Accession: S56445; G65233
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:95334362; PMID:7610040
A:Accession: S56445
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-577 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AA97116.1; PID:g537061
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Snao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65233
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-577 <BLAT>
A:Cross-references: GB:AE000493; GB:U00096; NID:g2367360; PIDN:AAC77177.1; PID:g1790666;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ytfM
A:Start codon: GTG

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

QY 74 LITQOOEVLDEKQGTFLAEAPONVKTMLRSKGYFSSKVS-----TEKDGYTV--HIT 127
Db 43 LSTIESDEVTDDR---FRARVDDAIRGLKALGYOPTIEFDLRPPPKGRQVLIKVT 99
QY 128 PGPRTKIANVGVAIIIGDILSDGNLAERYRNALENWQOPVGSDFDQDQSWENSKTSVLGAVT 187
Db 100 PGVPVLIGTDVVLRGARTD---KDYLK--LLDTPAIGTVLNGQDYENFKKSIATIAL 154
QY 188 RKGY-----PLAKLGNTRAAVNPOTATVDLNVVVDGSRPIAFGDFEITGTQRYPEIVSG 242
Db 155 RKGYDFSEFTKAQLG---IALGLHKAFWDID--YNSGERYRFGHVTFEQSQ-IRDEYLON 208
QY 243 LARFQPGTPYDLDLLDLDLQOALEONGHYSGASVQADPDRLOGDRV-PVKVSVTEVKRHL 301
Db 209 LVPFEGEYESKDLAELNRLRSATGNFNSVVAPOQDKARETKVLPGLTGVVSPRTENT 268
QY 302 ETGIRLDSYGLGGKIAYDYNFNKGYIGSVVMDMKYETTLAAGISOPRNYRGNWTS 361
Db 269 ETGVGYSTDVGPVRKATW-----KKPMNSY-----GHSITSTISAEQ-----TL 311
QY 243 LARFQPGTPYDLDLLDLDLQOALEONGHYSGASVQADPDRLOGDRV-PVKVSVTEVKRHL 301
Db 209 LVPFEGEYESKDLAELNRLRSATGNFNSVVAPOQDKARETKVLPGLTGVVSPRTENT 268
QY 302 ETGIRLDSYGLGGKIAYDYNFNKGYIGSVVMDMKYETTLAAGISOPRNYRGNWTS 361
Db 269 ETGVGYSTDVGPVRKATW-----KKPMNSY-----GHSITSTISAEQ-----TL 311
QY 362 NVSYNRSTONLEKRAFSGGIWYVDRAGIDARLGAELFAGRKIPGSDIDLGNSHATML 421
Db 312 DFSYKMPLLKN-----PLEQYLVQG--GFKRTDLNDESSTTL 349
QY 422 TAS-----WKROLLNNVLPENGHYLDGKIGTTLTGLTSLSTALIRTSARAGYF---- 469

Db 350 VASRYWDLSSGWRRAI-----NLRWSLDHFTQGEINTNTMLFYPGVMISRTSRGGLMPTWG 406
QY 470 -----FPPENKKLGT-----FIIRGOAGYTVARNADVPSSL 501
Db 407 DSORYSIDYSNTAWGSDYDFSVFQAQNVMIITLYDRIRFVTRGTLGNIETGDFDKVPPDL 466
QY 502 MFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSCGAVFHDMDGDA 561
Db 467 RFFACGDRSIRGYKYKSIAPKYANGDLKGASKLITGSLEYQYVNTGKWGAVFVDSGEAV 526
QY 562 ANFKRMKMLKHGSLGVRWFSPLAFSFDIAYGHSDKK---IRWHISLG 606
Db 527 SDIRKSDEFTGTGVGRWESPVGPIKLDFAVPVADKDEHGLQFYIGLG 574

RESULT 5

F91278
N:Alternate names: hypothetical protein ECs5198 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F91278
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A:Reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: F91278
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <HAY>
A:Cross-references: PIDN:BA38621.1; PID:g13364675; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs5198

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

QY 74 LITQOOEVLDEKQGTFLAEAPONVKTMLRSKGYFSSKVS-----TEKDGYTV--HIT 127
Db 43 LSTIESDEVTDDR---FRARVDDAIRGLKALGYOPTIEFDLRPPPKGRQVLIKVT 99
QY 128 PGPRTKIANVGVAIIIGDILSDGNLAERYRNALENWQOPVGSDFDQDQSWENSKTSVLGAVT 187
Db 100 PGVPVLIGTDVVLRGARTD---KDYLK--LLDTPAIGTVLNGQDYENFKKSIATIAL 154
QY 188 RKGY-----PLAKLGNTRAAVNPOTATVDLNVVVDGSRPIAFGDFEITGTQRYPEIVSG 242
Db 155 RKGYDFSEFTKAQLG---IALGLHKAFWDID--YNSGERYRFGHVTFEQSQ-IRDEYLON 208
QY 243 LARFQPGTPYDLDLLDLDLQOALEONGHYSGASVQADPDRLOGDRV-PVKVSVTEVKRHL 301
Db 209 LVPFEGEYESKDLAELNRLRSATGNFNSVVAPOQDKARETKVLPGLTGVVSPRTENT 268
QY 302 ETGIRLDSYGLGGKIAYDYNFNKGYIGSVVMDMKYETTLAAGISOPRNYRGNWTS 361
Db 269 ETGVGYSTDVGPVRKATW-----KKPMNSY-----GHSITSTISAEQ-----TL 311
QY 362 NVSYNRSTONLEKRAFSGGIWYVDRAGIDARLGAELFAGRKIPGSDIDLGNSHATML 421
Db 312 DFSYKMPLLKN-----PLEQYLVQG--GFKRTDLNDESSTTL 349
QY 422 TAS-----WKROLLNNVLPENGHYLDGKIGTTLTGLTSLSTALIRTSARAGYF---- 469
Db 350 VASRYWDLSSGWRRAI-----NLRWSLDHFTQGEINTNTMLFYPGVMISRTSRGGLMPTWG 406
QY 470 -----FPPENKKLGT-----FIIRGOAGYTVARNADVPSSL 501
Db 407 DSORYSIDYSNTAWGSDYDFSVFQAQNVMIITLYDRIRFVTRGTLGNIETGDFDKVPPDL 466
QY 502 MFRSGGASSVRGYELDSIGLAGPNSVLPERRALLVGSLEYQLPFTRTLSCGAVFHDMDGDA 561

Db 467 RFAGGDRSIRGKYKAPKANGDLGASKLITGSLSEYQYNNVTGKMWCAVFDSEAV 526
QY 562 ANFRMKLKHGSLGVRRWPSPLAPSFDAIGHSDKK---IRWHISLG 606
Db 527 SOIRRSDFKTGTGTVGVRRWSPGPKLDFAVPVADKDHGLQFYIGLG 574

RESULT 6
F86119
hypothetical protein ytfM [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F86119
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F86119
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <STO>
A:Cross-references: GB:AE005174; NID:gl21519219; PIDN:AAG59418.1; GSPOB:GN00145; UWGP:258
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ytfM

Query Match 12.0%; Score 379.5; DB 2; Length 577;
Best Local Similarity 24.7%; Pred. No. 1.2e-17;
Matches 145; Conservative 87; Mismatches 245; Indels 111; Gaps 20;

QY 74 LITQOOEVLDDKEQTGFLAEAPDNVKTMLRSKYSKSL---TEKDGAVTV--HIT 127
Db 43 LSTIESDEVTDDR---FRAYDDAIRGLKALGYQPTIEFDLRPPPKGRQVLIKAVT 99

QY 128 PGPRTKIANGVAILGDILSDNLAERYRNALENWQPVGSDFOODSWHNSKTSVLGAVT 187
Db 100 PGPVLTGGTDVYLRGARTD---KDYLK--LLDTPRAIPGVLNQGDYENFKSLTSIAL 154

QY 188 RKGY-----PLAKLGNRAAVNPDTATVLDNVVDSGRPTAFQDFEITGTQRYPEQIVSG 242
Db 155 RKGYDFSEFTKAQLG---IALGLHKAFWIDID--YNSGRYRFCHVTFEGSQ--IRDEVLQN 208

QY 243 LARFOCTPYDLDLLDFOQALPEONGHYSGASVQADFDRLQGRV-PVKYSVTEYKRHLK 301
Db 209 LVFPEKEDEYESKDLELNRRLSATGWFNSVVVAPQFDKARETKVLPRTGRVSPRTENTI 268

QY 302 ETGIRLDSEYGLGKTIAYDYNFLNKGYIGSVYVMDMKYETTLAAGISOPRNYRGNYWTS 361
Db 269 ETGVGYSTDVGPVKATM-----KKPMNSY-----CHSLTSTSIAPQ-----TL 311

QY 362 NVSYNSTONLEKRAFSGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATML 421
Db 312 DFTKMPLLKN-----PLEQYFLVQG-GFKRTDLNDESSTLL 349

QY 422 TAS-----WKROLLNVLPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYF--- 469
Db 350 VASRYWDLSSGWQRAI---NLRWSLDHFTQGEITNTTMLFYPGVMISRTSRGGLMPTWG 406

QY 470 -----FTPENKLGCT-----FIIRGOAGYTVARDNADVPSSL 501
Db 407 DSORYSIDYSNTANGSDVDFSVFQANVWIRTLVDRHRFVTRGTLGWIEGDFDKVPPDL 466

QY 502 MFRSGGASSVRGYELDSIGLAGNPSVLPERALLVGSLEYOLPRTTLTSGAVFDMGDAA 561
Db 467 RFAGGDRSIRGKYKSIAPKANGDLGASKLITGSLSEYQYNNVTGKMWCAVFDSEAV 526

QY 562 ANFRMKLKHGSLGVRRWPSPLAPSFDAIGHSDKK---IRWHISLG 606
Db 527 SOIRRSDFKTGTGTVGVRRWSPGPKLDFAVPVADKDHGLQFYIGLG 574

RESULT 7
B64012
hypothetical protein HI0698 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64012
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64012
A:Molecule type: DNA
A:Residues: 1-578 <TIG>
A:Cross-references: GB:U32752; GB:L42023; NID:gl573692; PIDN:AAC22357.1; PID:gl573700

Query Match 11.8%; Score 374; DB 2; Length 578;
Best Local Similarity 22.9%; Pred. No. 2.9e-17;
Matches 153; Conservative 93; Mismatches 263; Indels 158; Gaps 18;

QY 3 IKPTALLPALFFPHAYAPAADLSENKAAGFALFKKNSPOTESVKLKPFPVRIIDTQDS 62
Db 6 LKUTALFL-ALSCFPAPAEQTVDIEVQGIKGFRAVRNTD-----LNVLNLIK 51

QY 63 EIKDMVEEHLPLITOOQEEVLKDEQTFGLAEAPDNVKTMLRSKYSKSLVSTKED--- 119
Db 52 EEMDGSERYOHLVTKA-----VDRLRVFGYESSVRFERKORQG 91

QY 120 --GAYTVHTIPGPRTKIANVGVAIGDILSDNLAERYRNALENWQPVGSDFOODSWEN 177
Db 92 KRDLIIAHVTPGEPPTKIAGTDVQIEGAAQDENF-----NALRKNLPKDGVLVHEHOTYDD 146

QY 178 SKTSVLGAVTKRGYPLAKLGNTRAAVNPDTATVLDNVVDSGRPIAGDFEITGTQRYPE 237
Db 147 YKTAISKALNRYGPDGNFKISRLSEIPETHQAMRMILFDSGVRYHYGNTITFHSQ--IRD 205

QY 238 QIVSGLARFQPTPYDLDLLDFOQALPEONGHYSGASVQADFDRLQGRV-PVKYSVTEVK 237
Db 206 DYLNLTNLTKSGDPYLMNNLSDLTSDFPSSNWFSSVLVQPNVNH-KSKTVDVEIILYPRK 254

QY 298 RHKLEGTIRLDSYVGLGGKIAY----- 319
Db 265 KNAMELGVESTDGGVHGQIGTKPWINSRHSLSRNLVLSAPKQTLFATYRMPLLKNPL 324

QY 320 DYNNLFNKGYIGSVYVMDMKYETTLAAGISOPRNYRGNYWTS-----NVSYNRST 369
Db 325 NYTYDFAVGEKEKENDTNTRVLTLSA-----LRYMNNAHGWQYFGGLRMRYDSFT 375

QY 370 TQNLKRAF-----SGGIWYVRDRAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASW 425
Db 376 QADITDKTLLLYPTVGFTTRLRGGSFATWG-----DVQKI---TFDLs----- 416

QY 426 KRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALIRTSARAGYFPT-PENKKGTLGTPIIRG 484
Db 417 KR-----IWLSESSFILKVQASSAWVRYTAENHRV---VARA 449

QY 485 QAGYTVARDNADVPSSLMFRSGGASSVRGYELDSIGLAGNPSVLPERALLVGSLEYOLP 544
Db 450 EIGYLLHTKGTGTEKIPPTLRFPAAGDVSVRGYKKTAPKRNNGKLVGSRLLTTSLEYQYQ 509

QY 545 FTRTILSGAVFDMGDAAANFKRMKLVKHGSLGVRRWPSPLAPSFDAIGHSD-----KKIR 600
Db 510 VYPNWAATFADSLAADNYTAKELRYGTCVGRWASPVGAIKFDIATPIRDKNSKNIQ 569

QY 601 WHISLGT 607
Db 570 FYICLGCT 576

AE0428

probable exported protein YP03524 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AE0428
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.; Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, N.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AE0428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-578 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92753.1; PID:g15981448; GSPDB:GN00175
C:Genetics:
A:Gene: YP03524

```
Query Match      11.6%; Score 368.5; DB 2; Length 578;
Best Local Similarity 22.8%; Pred. No. 6.8e-17;
Matches 153; Conservative 93; Mismatches 264; Indels 161; Gaps 21;

QY 1 MWIKPTALLPALFFPHAYP-----AALISENKAAGFALFKKSPDTSVKLKP 51
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MLRYPILCFVCLLATPIAANVRLOVEGLSGLDILNVRRLSTI-----GTDEVTADG 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 52 KPVRIOTQDSEIKDMVEEHLPLITQQOEVLDKQGTGFLAEEADPNVKTMLRSGYFSS 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 56 RFRSRVD-----FAIROGLRALGYDYP 77
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 KVSITEKD-----GAVTVHTPGPKTKIANVGVAITGLDILSDGNLAEYRYNNALENMQOP 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 TITPELQNRPAAPARVLLIAKVPGEPLVAGVDIVLOGAKTDPD-----YQALVRRDTPK 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 166 VGSDFQDQSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDSSGRPIAFG 225
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 134 IGSVLNHGDFDNFTSSITGALARGYFDANMIKSQLGVAAQHEAFNIDFDSGORYRFG 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 DFEITGQRYPEQIVSGIARFQPGTPYDLDLLDFOQALEQNGHYSGASVQADF--DRLOG 284
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 194 KVIFQSGQ--IREDYLNQLVFPFHEGEFTSDELAELNRLAATNWSNVVSPDFQDAKES 252
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 285 DRVPKYSVTEVKRHKLETGIRLDSEYGLGGKIAYDYNFLNKGKIGVSGVWMDMXETTL 344
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 253 KILPLDAVTP-----RTENTVELGGYATD-----VG-----PRL 283
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 345 AAGISQPNRYRGNTWTSNVSYNRSTQNLEKRAFGSGGIWYVRDRAGIDARLGAEFL 400
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 284 TASWRKP-----WNSFGCHSLTTTALSAPQTLDFSYRIPLLRN-----PLEQYIL 330
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 401 AEG--RKIPGSDIDLGNSHATM-----LTASKRQLLNVLHPENGHYLDGKIGTTL 450
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 IOGGRFR--TDLNDTNSDTTLLNVARFMDLSSGMORAI---NLRWSLDHFTQGRVTDYT 384
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 451 GTFLSSTALIRTSARAGY-----FPTPEN---KKLG--- 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 385 MLLYPGVSYNTRGRCAMPVWGDQSORYSIDVSDTTWGSDDVDFGIFQAOQVWIRLTCEKN 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 479 TPIIRGQAGYTVARNADNPVSGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERRALLVGS 538
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 RPARVANGVWIEETNFRNPVPSLRFAGGDRSIRGYKFNITSIPRSDSKGLTGASKLATGS 504
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 539 LEYQIPFTRTLGSAVFHDMGDAANFKRMKLKHGSGLGVRWFSPLAFSFDIAYGHSDKK 598
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 505 LEYQVNTFTGRWGAFFVDSGEAVNNTSKSDLTGAGGVYRWASPVGPIKLDIAAPIGDNE 564
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 599 ---IRWHISLG 606
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 565 THGVQFYIGLG 575
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 9

AD1055

probable exported protein ytfM [imported] - Salmonella enterica subsp. enterica serov C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD1055
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Par, S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica ser A:Reference number: AB0502; PMID:11677608
A:Accession: AD1055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD06889.1; PID:g16505537; GSPDB:GN00176
C:Genetics:
A:Gene: ytfM

```
Query Match      11.2%; Score 355.5; DB 2; Length 577;
Best Local Similarity 22.3%; Pred. No. 5.1e-16;
Matches 134; Conservative 99; Mismatches 267; Indels 101; Gaps 17;

QY 55 VRIDTQDSEIKDMVEEHLPLITQQOEVLDKQGTGFLAEEADPNVKTMLRSGYFSSKVS 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 26 LKVEGLSGELENRYRAOLSTI--QSDEVTDDR---FRARVDDAIREGKALGYEPTIK 80
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 115 L-----TEKDGYTVHTPGPKTKIANVGVAITGLDILSDGNLAEYRYNNALENMQOPVGS 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 81 FDLPLPPPAKGRQVLIARVTPGPVLGGTVEILRGARTD---KDYL--ALLKTRPAIGT 135
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 169 DFDQDQSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVVDSSGRPIAFGDFE 228
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 136 VLNCGDYDNFKKSLTSLRKGYFDSEFTKSQLGIALGRHQAFWIDIDYDSGERYRFGPVT 195
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 ITGTRQRYPEQIVSGIARFQPGTPYDLDLLDFOQALEQNGHYSGASVQADFQRLQGRDV- 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 196 FEGSQ--IRDEVLQNLFPKEGDEYESKDLAELNRLSATGNFNSVVAPEPEKSRKTKIL 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 288 PVKYSVTEVKRHKLETGIRLDSEYGLGGKIAYDYNFLNKGKIGVSGVWMDMXETTLAAG 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 PLKGVVSPRTENTITGVGYSTDVGPVKASW-----KKPMWNSY-----CHSLTUTTS 303
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 ISQPNRYRGNTWTSNVSYNRSTQNLEKRAFGSGGIWYVRDRAGIDARLGAEPLAEG--RK 405
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 304 ISAPEQ-----VLDESYKMPLLKN-----PLEQYILVGGGPKR 336
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 406 IPGSDID-----LGNSHATMLTASWRQLLNVLHPENGHYLDGKIGTTLGTFLSSTALI 460
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 TDLNDEODSTTLAVSRYWDLSSGMORAI---NLRWSFDHFTQGRVNTTMLFYPGVMIS 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 461 RTSARAG-----YFPTPENKKLGT-----FIIRGQAGY 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 RTRSGGLMPTWGDQSORYSDVSNATWGSDDVDFSLQAQNVWIRLTLDHRRFVRNANLGW 453
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 489 TVARNADNPVSGLMFRSGGASSVRGYELDSIGLAGPNSGVLPERRALLVGSLEYQLPFRFT 548
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 454 IETGDFKVPDPLRFFAGGDRSIRGYKFNITSIPKSDSGNLKASKATGSLGLEYQYVNTGK 513
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 549 LSGAVFHDMDGDAANFKRMKLKHGSGLGVRWFSPLAFSFDIAYGHSDKK---IRWHISL 605
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 514 WMGAVFVDSGEAVSDIRKDFKTGTGVGVYRWASPVGPKVLDFAVPVGDKDDEHGLQFYIGL 573
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 606 G 606
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 574 G 574
```

RESULT 10
F82064
conserved hypothetical protein VC2548 [imported] - Vibrio cholerae (strain N16961 ser

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82064
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: F82064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <HEI>
A:Cross-references: GB:AE0033852; NID:g9657119; PIDN:AAF95689.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2548
A:Map position: 1

Query Match 11.08; Score 348; DB 2; Length 582;
Best Local Similarity 23.18; Pred. No. 1.6e-15;
Matches 136; Conservative 96; Mismatches 255; Indels 102; Gaps 19;

QY 65 KDWVEHLPLITQQ-----EEVLDEQTFGLAEAPDNVKTMLRSKGYFSSKVSLT- 116
DB 46 KDNVEYLLSIAQDYSLRFQSLERSMT-----EA-----LNALGYTHPSIDFTV 93

QY 117 -EKDGYTVHTPGPRTKIANVGVAILDLSQNLAEYRNALEMMQPVGSDFDQSDW 175
DB 94 SEDNORLRAAVTLGEVTRLSVDIVIRGEAGDRDFORLIRS-----GLRVDAPLNHSLY 149

QY 176 ENSKTVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVNVVDSGRPIAFDFTGTORY 235
DB 150 DNLKSGIRNALQKGYFNGDFOASRLVPELNOARVILHFDGRIYLFQATTVESQ-I 208

QY 236 PEQIVSGLARFOGPPYDLDLLDFOQALEONGHYSGASVQADFRL-QGDRVPVKVSVT 294
DB 209 DENRVMSLRFPGQGPYLVSGVGFENLNTDMFSSVFVEPDLSQLDQRELPLKVTILA 268

QY 295 EVKRHKLETRLDSEYGLGCKIAYDYNNFNKGYIGSVVMDMDKYETTLAAGISOP--- 351
DB 269 POARNOLETGLGYSTDVGRSLKWKPKWNSQGHSPDSFSLSIPEQITAGYKIPLD 328

QY 352 ---RNYRGNYWTSVNYNRSTQNLKRAKSGGLWYVDRDAGIDARLGAELAEGRKIPG 408
DB 329 ALNEYRQYQYGMKHL--DKRQTESLENSLSLERNH----- 361

QY 409 SDIDLGNSHATMLTASMKRQLLNW---LHPENGHY-LDGKIGCTTLGT----- 452
DB 362 -QLD-CGWHRTVFI---RYLLENYRQGLQDDNSQFLPCMTYTRRTYRSNSGLLTWGDK 415

QY 453 -----FLSTALIRTSARAGYFTT-PENKKLGTTIIRGQAGYTVARDNADVPFG 500
DB 416 QTITLEYCDPALLSETRYLRLOTGSSRLTYARNHRA---LVRVDGGANLVDFEQLSPS 472

QY 501 LMFPSGASSVRYELDSIGLAGNGSVLPERALLVGSLEYQLPFTRTLGSVAFVHDMGDA 560
DB 473 LRFAGGDNLRGYCKYSLSQDASGALTGAKYIATSSIEYQYELTGNWNAWAEVMDVGD 532

QY 561 AANFRMKLKHGSLGVNRFSPAPFSDTIAYG---HSDKKIRWHISLG 606
DB 533 FND--NPEKKKGVTGIRWISPVGPRLDFAWGLDAAAPGDFKIHFTLG 579

RESULT 11
H97672
hypothetical protein AGR_C_4742 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97672
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H97672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-641 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88337.1; PID:g15157817; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4742
A:Map position: circular chromosome

Query Match 10.28; Score 323.5; DB 2; Length 641;
Best Local Similarity 22.68; Pred. No. 8.4e-14;
Matches 148; Conservative 94; Mismatches 290; Indels 123; Gaps 21;

QY 6 TALLLPALFFPHAYAPAADLSENKAAGFALFNKSPDTSVKLKPKFPVRIDTQDSE--- 63
DB 19 TALAVALTALYPAP--ARDAPAFKLFQMLWSEPEVEVIN-PVKYATVTLDAADAKS 75

QY 64 IKDWVEEHLPLITQQEEVLDEKQTFGL-----AEEAPDNVKTMLRSKGYFSSKVSLT-- 116
DB 76 LKSSLESSLALLADK-----DKPASGDLGLLIKARDORRLIAALYENARYGGIVNVTVA 130

QY 117 -----EKOGAYTVHTPGPRTKIANVGVAILDLSQNLAEYRNALENN 161
DB 131 GKNVDDLPPNPVFDISTVPVMTVTTPGPFTLGN--VRLEGDV-TGRNLDEGLIA--- 184

QY 162 WOQPVGSDFDQSDWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLVNVVDSGRP 221
DB 185 -----GGDAGSLAIIRAGNKLIDDLKAEGRPALKLTKREAVANHAHTVTDITMAEGGPV 239

QY 222 IAFGOFELTGTORYPEQIVSGLARFQPCPTPYDLDLLDFOQALEONGHYSGASVQADFDR 281
DB 240 APLGAVTVTGEKTVGDDFIRYRSLNGGEPYSEPKLKAADRLQLGVFSSLTIKAGTIL 299

QY 282 LOGDRVPVKVSVTEYKRHKLETRLDSEYGLGCKIAYDYNNFNKGYIGSVVMDMDKYETTLAAGISOP--- 333
DB 300 ARDGTIPLTIEVSEGHRYFGVGAQYSTEGIGLQGYWGHNRNLFQGAESLRTEGVSRTA 359

QY 334 ---VWMDM-----KVETTLAGISQPNRYRGNYWTSVNYNRSTQNL 372
DB 360 EASSVEGMDYSAGITFTPGMPNPTFTKSLIAKTENPDYRAKTLGT----- 409

QY 373 LKRAFSGGIWTVRRDAGIDARLGAELAEGRKIPGSDIDLGNSHATMLTASMKRQLLNW 432
DB 410 -----AGFAYELND-----TDTAAGLEVQWADTEDAFGKNKYEYLTTSIPLEFVND 454

QY 433 VLHPENGHYLDGKIGCTTLGTFLSSTALIRTSARAGYFTTPEP-----KKLGT---FIIR 483
DB 455 T-----RDKLNPTEGFRASLAAPSYEALNGTFFSFECSITCYKLGAEDELIMA 506

QY 484 GQ--AGYTV-ARDNADVPGLMFRSGGASSVRYELDSIGLAGNGSVLPERALLVGSLE 540
DB 507 GKLSGVLVGGSDLDIPTTRRFFAGGGGSRVGYOYEISPYNAAGDGTGGRSYVGVSE 566

QY 541 YOLPFTRTLGSVAFVHDMG---DAAANFRMKLKHGSLGVNRFSPAPFSDTIAYG 591
DB 567 ARKVTDTDTIGLYPFDAGVSDGVTPDES--DIRAGAGILRYATPFGPLRLDVA 619

RESULT 12
AF2897
conserved hypothetical protein Atu2615 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2897
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gilllet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; MCCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kan
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AF2897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-617 <KUR>

A:Cross-references: GB:AE008688; PIDN:AAL43596.1; PID:g17741114; GSPDB:GMO0186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu2615

A:Map position: circular chromosome

Query Match 10.1%; Score 320; DB 2; Length 617;

Best Local Similarity 22.4%; Pred. No. 1.4e-13;

Matches 147; Conservative 92; Mismatches 286; Indels 130; Gaps 21;

QY 6 TALLPALFFPHAYAPADLSNKAAGFALEKFNKSPOTESVKLKPFPVPIVDQDSE-- 63

DB 2 TFLYPAF-----ARDAFAKFLGMLNGSEPEVEVIN-PYKAVTLDAADAKS 51

QY 64 IKDWVEHLPLITQQQBEVLDKQTCGL-----AEEAPDNVKTMLRSKGYFSSKVSLT-- 116

DB 52 LKSSLENSULLADK-----DKPASDGLGLIKARDRDLTAALYENARYGGIVNTVA 106

QY 117 -----EKDGAAYTVHTTPGPRKIANVGVAILEDLSGNLAEEYRNALEN 161

DB 107 GKNVDDLPPNPVEDHSTPVPWVITVTPGPKFTLGN--VRLEGDV-TGRNDEYGLIA-- 160

QY 162 WQPPVGSDFQDQSWENSKTSVLGAVTRKGYPLAKIGNTRAANVPDPTATVNLNVVDVSGRP 221

DB 161 -----GGDAGSLAIRAGNKLIDDLKAERPLAKLTKREAVANHATNPTVITMAEGGPV 215

QY 222 IAFGDFETGTQRYPEQIVSGLARFQPGTPYDLDDLDFQOALEONGHYSGASVQADFDR 281

DB 216 APLGAVTVGKTVDCGDFIRYSRLNGGEPYSPEKLRKAADRLROLGVFSSLTKEAGTL 275

QY 282 LQGRVPVKSVEVKRHKLETGIRLDSEYLGKGIADYNNLFPNG-----YIGSV----- 333

DB 276 ARDGTPLTIEVSEGRHRYFGVAQYSTTEGIGLOGYGHRLNLFQOAELESRSVSRIA 335

QY 334 ----VWMD-----KYETLLAAGISOPRNYRGNWTSNVSNTQNTON 372

DB 336 EASSVEGMDYSAGITFKPCMFNPRITFTKSLIAKTENPDTYRAKTLGT----- 385

QY 373 LEKRAFSGIHYVDRAGIDARLGAFLAEGRKIPGSDIDLGNHATMLTASRWKRLNN 432

DB 386 -----AGFAYELND-----TDTAAAGLEVQWADTEDAFGKNEYLTTSIPLEFVRD 430

QY 433 VLPENGHYLDGKIGTTLGTFLSTALIRTSARAGYFFTPEN-----KILGT---FTIR 483

DB 431 T-----RDDKLNPTEGFRASLAAKPSYEALNCTFFSSFEFSITGVKGLGAEDRLMA 482

QY 484 GQ--AGYTV-ARDNADVPGLMFRSGASVRYGELDLSGLAGPNSVLPERALLVGSLE 540

DB 483 OKLSGVLVGGSDLDQDPTTTRRFAGGGSVRYGYSQYEISPYNAAGDATGCRSYVGSVE 542

QY 541 YQLPFTRTLSCGAVFHDG-----DAAANFRMKLKHGSGLVGRWFSPPLAPFFSFDIA 591

DB 543 ARIKVTDTIGLVPPFDAGVSDGVTPDFS--DIRAGAGIGLRYATPFGPLRLDVA 595

RESULT 13

G82706

conserved hypothetical protein xfl1231 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: G82706

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; PMID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: G82706

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-617 <SIM>

A:Cross-references: GB:AE003957; GB:AE003849; NID:g9106207; PIDN:AAF84041.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carrier

as-Neto, E.; Docena, C.; El-Dorri, H.; Paciniani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.B.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XFL1231

Query Match 9.7%; Score 306.5; DB 2; Length 617;

Best Local Similarity 23.3%; Pred. No. 1.1e-12;

Matches 140; Conservative 89; Mismatches 232; Indels 139; Gaps 25;

QY 82 VLDKEQTGF----LAEEAPDNVKTMLRSKGYFSSKVSLSL---TEKDG-AYTVHTTPGPRTK 133

DB 79 VVGKEQGEFLELYLLSQAEFTQRALEPFYGYAPTIRIDAPRONDHITVTVYDKGEVPR 138

QY 134 IANYGVAILGDISGNLAEEYRNALENMQQPVGSDFDQDQSWENSKTSVLGAVTRKGYPL 193

DB 139 VROAHVAMTAAQD---HYLQRLDEDFKPLGEIFNHTYEASKVIRLRAERGVD 194

QY 194 AKLGNTRAAVNPDATVNLNVVDVSGR-----PIARG-DFEITGTQRYPEQIVSGLARFQ 247

DB 195 ADFTHRRVETRAAAHADIDLWESGRRYDMGVPVRPHYDY-----FHEGLNPLVYWD 247

QY 248 PGTPYDLDDLDFQOALEONGHYSGASVQADFRLQGD-RVPVKVSVTEVVRHKLKLEGTIR 306

DB 248 EGSYFHEGLDLRLRESLTKLDYFSSIDIQPKPEADPEGNVPVDVKLERAKSKIVTAGIS 307

QY 307 LDSEYLGKGIADYNNLFPNGKYGVSVDMDKYEYTLAAGISQRP-RNYRGNWTSNVS 365

DB 308 YGSESGAG-----LRAGYRRYMRNAGHKMMARLDY 338

QY 366 NRSTQNLEKRAFSGGI-----WYVRDRAGIDARLGAELAEGRKIPGSDIDLG--- 414

DB 339 ----AQNLSLTTAYQYPAFKWLDGWYI-----FFARAYDEOTKYIDLNRVK 381

QY 415 -----NSHAT-----MLTASHMKROL-----LNNVLEPE-NGHYLDGKIGT 448

DB 382 LSAARSGQINRHLTATASLNLALGRWRYRADDGNTNTVAYQQSTLVYLPQLEASIVDDDAT 441

QY 449 TLTGFLSLTALIRTSARA---GYFFTPENKKLGTF-----IIRGQAGYTVARDNAD 496

DB 442 FPRNGSAATVLLRGASALGSKSNFTQLHGRLWFHGLGASSRLILRGAEAGTWTASDLVA 501

QY 497 VPSGLMFRSGGASVRYGELDLSIG--LAGPNSVLPERALLVGSLEYQL-----PFTRTL 549

DB 502 MPPLSRFFAGGVNSIRGYAFREVPRTAKPDALFALGAKHVFSGAGAEFFYYKGGFF----- 557

QY 550 SGAVFHDGDAANFKRMKLKH-GSGLVGRWFSPPLAPFFSFDIAYG--HSDKKIRWHISLG 606

DB 558 GGAVFVDSGSA---FNRYPDWHTGVGIGLRYRSPVGVPRVDIARGLNSPDSKIQLYIDIG 614

RESULT 14

A13488

outer membrane protein BME1895 [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: A13488

R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov

.. Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitens*
A:Reference number: AD3252; PMID:11756688

A:Accession: AF2089

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-623 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL53076.1; PID:gl7983939; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11895

A:Map position: 1

Query Match 9.2%; Score 290.5; DB 2; Length 623;

Best Local Similarity 20.4%; Pred. No. 1.3e-11;

Matches 140; Conservative 106; Mismatches 248; Indels 193; Gaps 28;

QY 38 KKKSPDTSVVKPK-PPVRT-----DTODSEIKOMVEEHLPLITQOOEEVLDKEQ 87

DB 15 KKQDPDI----IDPKTSVDVTTGDRKNADGKEADLKSVEGASGLVSDADKPA--SGS 68

QY 88 TGLAEAPD--NVKMLRSKGYESSKVS-----TE--KCGAVTVHITGCP 130

DB 69 AGLAKARGDYRRILSALYGRYGGTISIKVDGREANDIPDTEIPNNKVAITVDGCP 128

QY 131 -----RTKIANV-----GVAILGILSDGNLAEYRNALENWQPVGS 168

DB 129 QFLSRTAISNIAPPPGNNRRDKVOTPEENGFA--PGQEAKSQGTILKAERLAVENRQ---- 183

QY 169 DFDQDSWENSKTSVLGAVTRKGYLAKLGNTRAANVPDATTVDLNVVDSGRPIAFGDFE 228

DB 184 -----EGYAKARVTGEDVVDHADNRVSADIALDPGRKAYGPVS 223

QY 229 ITGTORYPEQIVSGLARFQCTPYDLDLLDFQALEQNGHYS--GASVQADFRLQGD- 285

DB 224 VVGTARMQPVQVAMWTLKPGQYDDP---DIENAKKRLGRMEVFRAMTFEADKIEPDG 280

QY 286 RVPKVSYTEVKRHKLETGIRLSEYGLGKGIAYDYNLFNKGY-----IGSV-- 333

DB 281 SLPIPLNQERKPRFGFGBYSIDFGVTSYVMHRLNGRERLREDAKVSIGGSDQ 340

QY 334 -VWMDKYETTLAAGISQPNRYGN-----YWTSNVSNRSTTQ----- 371

DB 341 NSFDPKNTYLLGASFAKPGVYTPDFTVATLDAKREVLDATETSIINAKTGTQIFDSE 400

QY 372 -----NLEKRAFSGGIWYVRD--RAGIDARLGAFLAEGRKIPGSDIDLGNSHATMLT 422

DB 401 LSGALYANASQCHEFVDVFGKRDPTTAGLEGNL-----LYDSR----- 438

QY 423 ASMKRQLLNVLHPENGHYLDGKIGT---TLGTFELSSALIRTSARAGYFETPENKILG 478

DB 439 -----NNKPDSSGFYLGNIOPIYEHYGNFATR-----FTAGERTYH 477

QY 479 TF-----IIRG--QAGTYVARDNADVPGLMFRSGCASSVRGYELDSIGLAGPNGSYLP 530

DB 478 GFGQDRLVVLACRLKVSIGVGSITADLPSPQLFAGGGSVRGYGRNIGVSAGNIIIG 537

QY 531 BRALLVGSLEYQLPTRTLSCGVFHDWG---DAANFKRMKLKHGSGLGVRWFSPLAPF 586

DB 538 GRSLSVEANGEVTRITDSIGNAVFVADAGYGEKSFDFDSE-QNRVGVGGGLRYLTSLGPI 596

QY 587 SPDIAV-----GHSOKKIRWHISLQTRF 609

DB 597 RLDAVPLNRRSGDPNFGYVYGQAF 623

RESULT 15

AF2089

hypothetical protein alr2269 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AF2089
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriugu
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2089

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-833 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA073968.1; PID:gl7131360; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2269

Query Match 8.8%; Score 279; DB 2; Length 833;

Best Local Similarity 22.9%; Pred. No. 1.2e-10;

Matches 153; Conservative 82; Mismatches 234; Indels 200; Gaps 26;

QY 81 EYLDKEQGTGLAEAPDNVKTMLRSK-----GYFSSKVSUTEKDG 120

DB 225 EYLVRPQSGQLTPELETQVYVIRTQPGRTTTRSQLQEDINALFGTGFTSN-VQASPEDT 283

QY 121 AYTIVHT---PGP---RTKI-ANVG-----AILGDLSDGNLAEYR 156

DB 284 PLGVRVSIVQNPVLSKVEIQANPGTNVPSVLPQATADEIFRAQYKILNLRDQEGIK 343

QY 157 NALENMQOPVSGDFDQDSWENSKTSVLGAVTRKGYPLAK-LGNTRAANVPDATTVDLNV- 214

DB 344 ELTKRYQD-----QCYVLNAVVGAPVSEN---GVVTLQVA 376

QY 215 -----VDSGRPIAFGDFEITGTORYPEQIVSGLARFQCTPYDLDL 256

DB 377 EGVVENISVFRFNKEGQDVNQGP-----RGRTQDYIITREVLKPGQVFNRT 427

QY 257 LLDFQOALEONGHYSVASVQADFRLQGDRLVPKVSYTEVKRHKLETGIRLSEYGLGK 316

DB 428 VOKDLQRVFGTGLFEDVNVSLD-PGTDPTKVVNVVNVRSRSGSIAAGAGISSSSQLFCT 486

QY 317 IAYDYNL-----FNKGYIGSVVWMDKYETTLAAGISQPN 353

DB 487 VSYQQONLNGRNQKLGAEVOLGERELFLDLRFTDPWLGG-----DPYRTSYTANIFRRS 541

QY 354 YR-----GNYWTSN-----VSYNRSTTONLEKRA---FSGGI 382

DB 542 ISLIFDGDDEIHTDPGNPNDTNGQDRPRVTRLGGGVTFTRPLSANPFEAEWTFASGL 601

QY 383 WY-----VRDRAGIDARLGAFLAEGRK-----IPGSDIDLGNSHATMLTASMKRQLLNVL 434

DB 602 QYRVSTRDADGNLNRKDGAVFDNGNRTSEIVPLSFSGTGEDDLLVLQVLAQDRLRNPL 661

QY 435 HPENGHYLDGKIGTTLGTFLLSSTALIRTSARAGYF-----FTPENKKLGTFIIRGQAGY 488

DB 662 QPTSGFL--RFGVDQSVPGSGCNIFLTRFRGSYSQYLPVKFTFGSKGPETIAFNIQGT 719

QY 489 TVARDNADVPGLMFRSGCASSVRGYELDSIGLAGPNGSVLPREALLVGSLEYQLPFTTR 548

DB 720 VL---GDLPPYEATLGGSNSVRGYEGALGSG-----RSFQVQASVEYRPPVSV 766

QY 549 LSGAVFHDMD-----AAANFKRMKLKHGSGLGCVLWVWSPPLAPFSDIAYG-HSDKKIR 600

DB 767 VSGALFFDVGSDLGSTRTTAELVNLKSGSGYGLGVRVQSPGLGPIRID--YGINDDGDGR 824

QY 601 WHISLQTRF 609

DB 825 INFGIGERF 833

RESULT 16

B87448

conserved hypothetical protein Cci1603 [imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

A:Residues: 1-797 <TEXT>
A:Cross-references: GB:AEO00375; GB:AEO02098; NID:g7225394; PIDN:AAF40639.1; PID:g722540
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0182
C:Superfamily: protective surface antigen D-15

Query Match 8.0%; Score 253; DB 2; Length 797;
Best Local Similarity 21.9%; Pred. No. 6.4e+09;
Matches 135; Conservative 89; Mismatches 248; Indels 144; Gaps 26;

QY 97 DNYKTLRSGKYF-----SSKVSTEKDGAVTTHIT--PGPRTKIANVGVAAILGDILSDG 149
| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 223 EKVTDFYQNNGYVDFRIILDIDIQNKDTKTIKITIVHEGGRFRWCKVSI-----EG 274
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 150 NLAEYYRNALE-----NQOOPVGSDFDODSHENSKTSVLGAVTRK-----GYPLAKLG 197
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 275 DTNEVPRAELEKLLPMKCGKWYE-----ROOMTAVLGEIQNRMSGAGAYSET- 322
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 198 NTRAAYNPDTATLVNVVDVDSGRPIAFDCEFTIGTORYPEQIVSGLAREFPQTTPYDLIDL 257
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 323 SVOPLPNAETKTVDVFLHIEPGRKIYVNEIHTGNKNKTRDEVVRRELQWESAPYDTSKL 382
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 258 LDPQQAALQONGHYSGASVOADFRLQG--DRVPKVSVTEVKRHKLETGIRLDSEYGLGG 315
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 383 QRSKERVELLGFDD--NQOFDAVPLAGTPDKVDLNMSLTERSTGSLDISLAGWVQDTGLVM 440
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 316 KIADYINLFNKGYGIGSVVMDMKYETTL-----AAGISOPNRYRGNTW---- 359
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 441 SAGVSODNIFFGTG--KSAALARSRSKTTLNGSLSFDPYFTADGVSLGYDVYKAPDPRK 498
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 360 -TSNYSYNRSTTONLEKRAFSGGIWYVRDRAGIDA-----RLGAEFLEAGRKI 406
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 499 ASTSIKOYKTTTAGACIRMSVPVTEYDRVNFGLVAEHLTVNTYNNKAPKHAYDFIKYKGT 558
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 407 PGSDIDLGNATHML--TASWKROLLNNVLPENGHYLOGKIGCTTIGTFLSSTALIRTS A 464
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 559 DCTD--GSFGWLXKGTGVGNRNKTDLSALMPTRG-YLTG-----VNAAELPGS 604
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 465 RAGYFTFPNK-----KLGTFIIRGOAGTVARDNA-DVPSGLMFRSGGASSGYVEL 516
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 605 KLOYYSANTHNQTFPPSLKSTFTLMLOGEVGLAGGYGTFKEIPFFENFYGGGLCSVRGE- 663
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 517 DSIGLAGPN-----GSVLV---BRALLVSGLEYQLP---FTRTLSGAVFHDMGD----- 559
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 664 --SGTLGPKVYDEYGEKITSYGGNKANVASIELLFPMPGAKDARTVRLSLFADAGSWWDGK 721
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 560 -----AAANFKRM-----KLKHGSGIGVRFSPPLAFPSFDIAY----G 593
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 722 TYDDNSSSATGCRVONIYGAGNTHKSTFTNELRYSAGGAVTWLSPLGPNKFYSVAYPLKK 781
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 594 HSDKKIRWHISLGRFP 609
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 782 PEDETQRFOFQLGTTTF 797
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 19
C70412
outer membrane protein - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: C70412
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; UID:58156666; PMID:9537320
A:Accession: C70412
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <AOF>
A:Cross-references: GB:AF000733; NID:g2983720; PIDN:AAC07292.1; PID:g2983730; GB:AE000675
A:Experimental source: strain VF5

[illegible]

RESULT 20

AC0129

probable surface antigen YPO1052 [Imported] - *Yersinia pestis* (strain C092)
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Park
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.;
Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Residues: 1-795 <KUR>

A: Cross-references. CR.

Genetics.

Genetics:

A; Gene: YPO1052

C; Superfamily: protei

Query Match 7.8%; Score 246; DB 2; Length 795;
Best Local Similarity 21.2%; Pred. No. 1.9e-08;
Matches 142; Conservative 105; Mismatches 273; Indels 150; Gaps 30;

QY 47 VKLKKPFVRIDTQDSIKQVMEHPLITQOOEVLVDKEQGTGFLAEAP----- 96
DB 163 VDLKLVFTGVSARIQI-NIVGNH-----SFTTDELISRFQ---LRDEVPMMNVVGDCKY 214

QY 97 -----DNVKTMLRSKGVF-----SSKVSUT-EKDGAY-TVHITPCPRTKIANGVA 140
DB 215 QKQKLAGDLETRSFYLDRGVARENDSTQVSLTPDKKIYVITINITEGPOFK----- 267

QY 141 ILGDILSDGNLAERYRNALENMOOPVGSDFDODSWENSKTSVLGAVTRKGYPLAKLGNTR 200
DB 268 -LNSIVISCNLAGHQAERLTKTEPGLFNGSKVTRMEDDIKKMLGRGYAYPRV-VTQ 325

QY 201 AAVNPDTATVLDNVVDSGRPIAFGDFEITGTORYPEQIVSGLARFORCPTPYDLDLDDLF 260
DB 326 PEINDDDTKVHLINVDAGNFYVRHIREGNDTSKDSVLRREMOMEGANLGNDOVEAG 385

QY 261 QOALBQNGHYSGASVQADFDRLQ--DRVPVKVSVTEVKRHKLETGIRLDSYGLGKIA 318
DB 386 KERLNLRYGFE--TVDVETORVPGAADLVDTYKVKERNVTGSLNFGIGYGTESGVFOVG 443

QY 319 YDYNNLKNKGIYSVWDMKYEITTLAAGISQPRNYRGNYWT-----SNVSYNRSTTQ 371
DB 444 VOODNWLGTGTVGINKNDYQTYAEFTLMDP-----YFTVGVSLGGRIFYNDFRAD 497

QY 372 NLEKRAFSGGIYVRDRAGIDARLGAEEFLAEGRKIPGSDIDLGNSH---ATML--TASWK 426
DB 498 NAD---LSG---YTNSSYVGADGTGLGF-INENNSL---RVGVYVHNDLSMDLPQVAMWR 547

QY 427 ROLLANVLHPNGHYLDGKIGTTLGTPLSSTALIRTSARAGY----- 469
DB 548 --YLESV--GERPGY-DGREGFTDDFTLNLGWTYNNLDRGFFPGSVKYSVNYKITVPG 602

QY 470 -----FTPENKKGITFII-RGOAGYTVARNADVPGLMRSGGASSVRGY 514
DB 603 SDNEFYKVTFTDSAYQPLNEDRSWVLGRGLGYDGIGSKEMPFYFAGGSGTVRGE 662

QY 515 ELDSIGLAG--PNC-----SVLPERALLVGSLEYOLP-----FTRLSCGAVPHD 556
DB 663 RSNNGIPRAAYANGATVNTSDAVGNAMAVASIELTTPPFISEKYSNVRTSIID 722

QY 557 MGD-----RANF-----KRMKLKHGSLGVRFWSPLAPSFDIAYG-----HSD 596
DB 723 SCTVVDNTNMENTAKTRAAGIDPYGKASNIRVSAGVALQWMSPLGLPLVFSYAKPVKDYEGD 782

QY 597 KKIRWHISLG 606
DB 783 KSEQFQFNIG 792

RESULT 21
B82099
Surface antigen VC2252 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82099
R:Heidelberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: B82035; MUID:20406833; PMID:10952301
A:Accession: B82099
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HEI>
A:Cross-references: GB:AF004297; GB:AE003852; NID:g9656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2252

A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 7.2%; Score 227; DB 2; Length 803;
Best Local Similarity 21.7%; Pred. No. 3.6e-07;
Matches 135; Conservative 95; Mismatches 225; Indels 168; Gaps 31;

QY 84 DKEOTGFLEAEAPNVKTMRLSKGYSFSSKVSUT-----EKDGAY-TVHITPCPRTKIAN 136
DB 212 DXYKKQVLAGDI-BALRTYYLDRGYLFQVDSQVAISPKKGVITLNLNAGEPYTVSK 270

QY 137 VGVATILGILSDGNLAERYRNALENMOOPVGSDFDODSWENSKTSVLGAVTRKGYPLAKL 196
DB 271 --VQPRGELM--GREAEF--TSLIPF--IGETINGSVTRLEESYKVKLGESGYAYPQV 322

QY 197 GNTRAAVNPDTATVLDNVVDSGRPIAFGDFEITGTORYPEQIV----- 240
DB 323 -RTIPEFDEKQOQVSLVHVVEAGKRVYRDIRFVGNNSTRDEVLRREMOMEGSWLNSKD 381

QY 241 --SGLARFQ-----PGTPYDLDLDDFOQALEON-----GHYSGASVQ 276
DB 382 IETGKTRNLRLGFEFTEVQTVRVPGSEDOVDLVYSVKEANSNVNFGVGYGTESGVSPQ 441

QY 277 ADFDR---LOGDRVPVKVSVTEVKRHKLETGIRLDSY-----GLGKKTAYDY 322
DB 442 VGLQDNFLGSGNRVGVNMIYQKN-----LTLEYRDPYNNLDVSLGKVV---FY 491

QY 323 NLFNKGYSVWDMKYEITTLAAGISQPRNYRGNYWTSNVSYNRSTTQNLKRAPSGGI 382
DB 492 NOFEASEAGIYDVTNYESGTSLTWGF--PFD-----ELNRFERGI 529

QY 383 WYVRDRAG-IDARLGAEEFLAEGRKIPGSDIDLGNSHAT---MLTASMKRQLNVLNHP 437
DB 530 GYTHNKIGNLTPYLOVENFLA---AQASNIDSGNLLTDDFDINLSWTRNNLNNSYPT 585

QY 438 NGHYLDGKIGTTLGTPLSSTALIRTSARAGYFTPENKKGITFIIIRGOAGY-----TVA 491
DB 586 AGNHORAEFYKMTVPG--SDAQYFKLOYDVRQYFPLTKKHEFTLLRGLRGLGNGYGQTDG 643

QY 492 RQNAOVPGLFRSGGASSVRGYELDSIGLAGP-----NGS-----VLPERAL 534
DB 644 KDNL-FPYENFYAGGFTSLRGFGSNS--AGPKAYRDYSGSNNGSDTATDSDVGGNAI 699

QY 535 LVGSLEYOLP-----PTRLSCGAVPHDMG---DAAANFKRMKLKHGS----- 573
DB 700 ALASVELIVPTPFASEEARNQIRTSIFYDMAVMDTEFDY-RGKADYGNQYTYDSDPTN 758

QY 574 -----GLGVRFWSPLAPSFEDIA 591
DB 759 YRSSYGVALQWSPMGPLVFSIA 781

RESULT 22
AB0530
outer membrane protein precursor yaeT [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0530
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church-
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far-
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; O'Neill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:g16501505; GSPDB:GN00176
C:Genetics:
A:Gene: yaeT

C:Superfamily: protective surface antigen D-15

Query Match 7.1%; Score 226; DB 2; Length 803;
Best Local Similarity 20.1%; Pred. No. 4.2e-07;
Matches 136; Conservative 102; Mismatches 284; Indels 154; Gaps 25;

Qy 47 VLKPKPPVRIDQDSEIKDMVEEHLPLITQOQEEVDKQETGLAEAP-----96
Db 163 VDLKLVFOEGVSAKIOOI-NIVGNH-----AFSTEELISHFO-----LRDEVPWNVGDRKY 214
Qy 97 -----DNVKTMLRSKGYF-----SSKVSLLT-EKDGAY-TVHTTPGPRTKIANVGVA 140
Db 215 QOKKLADLLETLSYLLDGYARFNIDSTQVSLTPDKKGIYTVNITEGQYKLSGVQVS 274
Qy 141 ILGDILSDGNLAEBYRNALNNQOVPVSDPDQSWENSKTSVLGAVTRKGYPLAKLGNT 200
Db 275 -----GNLAGHSAEIEKLFKIPGELYNGTKVTKMEDDIKLLGRGYAYPRV-QSQ 325
Qy 201 AAVNPDTATDLNNVVDSCRPIAFGDFEITGTQRYPEQIVSGSLARFPQGPYDLDLDD 260
Db 326 PEINDADKTVKLVNVVDAGNRFTYVKIRFEGNDTSKSVLRREMRQMEGAWLSDLDV 385
Qy 261 QOALEONGHYSGASVOADFDRLQ--DRVPVKVSVTEVKRHKLETGIRLSEYGLGKIA 318
Db 386 KERLNLRGFFE--TVDTDRQVPGSPDQVYVYKVKERNTGSENFGLGYGTSGVSFOAG 443
Qy 319 YDYNLNFNKGIGYGVVMDKYEITLAAGISQPRNRYGNWT-----360
Db 444 VODNWLGTGYSVINGTKNDYQTVSELSVTP-----YFTVGVSLGRIYNDPEAD 497
Qy 361 -----SNVYSNRSTQNL-----EKRAFSGGIWYVRDR-AGIDARLCA-EFLAEGRKIPGSDIDLGN 409
Db 498 DADLSOYTNKSYCTDVTGLFPINEYNTLRAGLVYVHNKLSNMQPOIAMDRYLE-----550
Qy 410 DIDLGNSHATMLTA-----SMKRLQNLNNVLPENH--YLDCKI-----GTTLTGTLSTAL 460
Db 551 --SKGPDASDAADDFPNYGNWTKNLDGCFYPTDGSRVNLTKV-TIPG-----SNEY 504
Qy 461 RTSARAGYFTFPENKKGTLFIIRQAGYTVARNADVPGLMFRSGGASSVRGYELDSIG 520
Db 605 KVSLDATATVPIDNDRKHWYVLTGRWYGDGLGKGEKMPFENFYAGSGSIVRGFSQNTIG 664
Qy 521 --LAGNGS-----VLPERALLVGSLEYOLP-----FTRT 548
Db 665 PKAVYKNGAHTSWDDDDYEDCTQESCKSDDAVGCNMAVASLEFIPFPFISEKYANS 724
Qy 549 LSGAVFHDWG-----DAAANFRMKLKHGSGLVGRWFSPLAPFSFDIAYG- 593
Db 725 VRTSFFWDMGTWMDTNDPSSAPSDVPDYSQGNIRMSAGIALQWMSPLGLVFSYAQPF 784
Qy 594 ---HSDKKIRWHISLG 606
Db 785 KYDGDGKAEQFQENIG 800

RESULT 23
A64742
hypothetical protein b0177 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64742
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.: Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64742
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-810 <BLAT>
A:Cross-references: GB:U000096; NID:g1786370; PIDN:AAC73288.1; PID:g1786374;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: protective surface antigen D-15

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;

Qy 97 DNVKTMLRSKGYF-----SSKVSLLT-EKDGAY-TVHTTPGPRTKIANVGVAAILGDLSD 149
Db 224 ETLRSYVLDGRYARFNIDSTQVSLTPDKKGIYTVNITEGQYKLSGVEVS-----G 275
Qy 150 NLAEBYRNALNNQOVPVSDPDQSWENSKTSVLGAVTRKGYPLAKLGNTRAAVNPDPAT 209
Db 276 NLAGHSAEIEKLFKIPGELYNGTKVTKMEDDIKLLGRGYAYPRV-QSMPEINDADKT 334
Qy 210 VDLNNVVDSCRPIAFGDFEITGTQRYPEQIVSGSLARFPQGPYDLDLDDLLFOQALQNGH 269
Db 335 VKLVNVVDAGNRFTYVKIRFEGNDTSKDAVLRREMRQMEGAWLSDLDVQDGKERLNLRG 394
Qy 270 YSGASVOADFDRLQ--DRVPVKVSVTEVKRHKLETGIRLSEYGLGKIAIDYNNLPNK 327
Db 395 FE--TVDTDRQVPGSPDQVYVYKVKERNTGSENFGLGYGTSGVSFOAGVQODNWLGT 452
Qy 328 YIGSVVWMDKYEITLAAGISQPRNRYGNWT-----SN 362
Db 453 GYAVGNGTKNDYQTVSELSVTP-----YFTVGVSLGRIYNDPEADADLSDYN 506
Qy 363 VSYNRSTQNL-----EKRAFSGGIWYVRDR-AGIDARLCA-EFLAEGRKIPGSDIDLGN 417
Db 507 KSTGTDVTLGFPINEYNTLRAGLVYVHNKLSNMQPOVAMWRYLYSMGEHP-STSDQDN 565
Qy 418 AT---MLTASWKRLNNVLPENH--YLDCKI-----GTTLTGTLSTALIRT 462
Db 566 KTDDETFNFGWTKNLDGCFYPTDGSRVNLTKVTPGSDNEYKVKVLT-----615
Qy 463 SARAGYFTFPENKKGTLFIIRQAGYTVARNADVPGLMFRSGGASSVRGYELDSIG 520
Db 616 ---ATVVPIDDDHK---WVLTGRWYGDGLGKGEKMPFENFYAGSGSIVRGFSQNTI- 668
Qy 521 LAGNGSVLPER-----ALLVGSLEYOLP-----544
Db 669 --GPKAVYVPHOASNYDPPDYECATQDQAKDLCKSDDAVGCNMAVASLEFIPFPFIS 726
Qy 545 --FRTLSCAVFHDWGDA-AAANFRMK-----LKHGSGLVGRWFSPLAPFSFD 589
Db 727 DKYANSVRSFFWDMGTWMDTNDPSSAPSDVPDYSQGNIRMSAGIALQWMSPLGLVFS 786
Qy 590 IAYG----HSDKKIRWHISLG 606
Db 787 YAQPFKKYDGDGKAEQFQENIG 807

RESULT 24

C90651
hypothetical protein ECs0179 [Imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90651
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90651
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <HAY>
A:Cross-references: PIDN:BA033602.1; PID:g13359635; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs0179
C:Superfamily: protective surface antigen D-15

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;

25:

[illegible]

```

RESULT 25
C85502
hypothetical protein yaeT [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85502
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose,
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoumis
Nature 409, 529-533, 2001.
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: 885480; MUID:21074935; PMID:11206551
A:Accession: C85502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <S>O
A:Cross-references: GB:AE005174; NID:g12512902; PIDN:AAG54479.1; GSPDB:GN00
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yaeT
C:Superfamily: protective surface antigen D-15

```

Query Match 6.8%; Score 217; DB 2; Length 810;
Best Local Similarity 20.5%; Pred. No. 1.7e-06;
Matches 127; Conservative 93; Mismatches 253; Indels 148; Gaps 25;
Qy 97 DNYKTMLRSGYVF-----SSKYSILT-EKDGAY-TVHTPGPTKIANNGVAILGDSG 149

[illegible]

```

Query Match      6.8%; Score 215.5; DB 2; Length 737;
Best Local Similarity 20.0%; Pred. No. 21e-06;
Matches 133; Conservative 83; Mismatches 240; Indels 209; Gaps

QY 105 SKGYFSKSVLTKDCAGTCAVHTTGPRTKTA-----VAILGD-IILSDGNL 151
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 142 ACGRYSAEIN-----AEVITPPNRVALKININECTVAAISHINVGVTFEEDL 192

```

Qy	152	ABYYRNALENMOQPVGSDFOQDQSWENSKTSLVGAVTR-----KGYPKALGNTRAAVN	204
Db	193	TDLFELKTTNMLSPFKND--DKYAREKLS--CDLERLSYYLDGYNINMDIASTOVST	247
Qy	205	PDATATVDLNVVDSRPTAGDFEITGQRYPEQIVSGLARFQG-----TPYDL-	254
Db	248	POKKHHVITVINEGEKTYTRDKLTGDLKVPPEEVKRLLLVQKGOVFSRKVMYTTSDLI	307
Qy	255	-----	254
Db	308	TRRLNGEYTFANVNGVPEAHDDDKTYSVTFFVDPCKRAYVNRINFRGNKTKTEDEVLR	367
Qy	255	-----DLLLDFOQA--LEQNGHYSICASQAQDFDLQDGRVPKAVSVTEVKRHKLET	303
Db	368	MROMEKGWASTYLLDQSKARLERLGTFFKEVNVETPAVPGTDDQDVNVYSEEPQSGITA	427
Qy	304	GIRLSEYV--LGCKIAYDYNLFNKGYIGSVWMDKYEITTLAAGISQPNRYGNWYT-	360
Db	428	SVGFAQSAGLILGGSISQN--NFLGTGNKVSICLRSEYTRYNFGFVDP-----YMTV	479
Qy	361	-----SNVSYNRSTTONLEKRAFSGGIWVDRAGIDARUG-----AEFLAEGRKIPGS	409
Db	480	DGVSLGYNAFYRKTYDDELVDVYAS---YSVNSLGGAKMSIGYPISETSLTYTGLSVQRD	535
Qy	410	DIDLG-----NSHATMLTAS--WKROLLNVLPBNGHYLDGKTGTTL-CT	452
Db	536	QIDTGRYVDEIYDFLQEGDNFTFKASIGNSESTLNKGLVATRGHSQSLLTLETTLPGS	595
Qy	453	FLSSTALIRTSARAGYFPTPEKKLGTFTIRQAGYTYVARDNAD--VPSGLMFRSGGASSV	511
Db	596	DLS--PYFKIDYR--GVQFAPLTONY--TMRFHTELCYGDGYGSTERLFPYENYVAGGNSV	650
Qy	512	RGVELDSIG-----LAGNGSVL-----PE-----RALLVGSLE--Y	541
Db	651	RGFKUSTILGPBSTSVARNPDGTPMKNGQPSDKGYHTDPODPAFGNGNIIITGGALLFL	710
Qy	542	QLPFT---RTLSGAVFHDMDGA-----AANFKRMKLKHGSGLCGRVWFSPLAFP	586
Db	711	PLPFPQDQORLTVLFDWDVGSTEDTDCPTKTTTNCDDGIKTDNLASSVGVGLTWTALGPL	770
Qy	587	SFDIA	591
Db	771	SESIA	775

RESULT 27
AE2417
hypothetical protein alr4893 [Imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2417
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 203-213, 2001
A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2417
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:BA000019; PID:BA000019; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4893

Query Match 6.6% Score 211; DB 2; Length 676;
Best Local Similarity 20.9% Pred. No. 3 3e-06;
Matches 132; Conservative 85; Mismatches 248; Indels 166; Gaps 27;
QY 76 TQQQEVLDD---KEOTGFLAEAPDNVYKTLR-----SKGYFSSKSLTEKDAYTVH 125
| : : : | | | : : : | : : | : | : |

Db 115 TPQEILRIQVKTGTG-----GDTSTQQLQRDVAAILETGLFPAS-----ANVNSR 160
 QY 126 ITPGPRTKIANVGVAILGDILSDCNLAEEYRNALENWQOPGVGDFDQDQSWENSKTSVILGA 185
 Db 161 TTPSCLNVVYQVPVIVRSQLTGAKALTYSAQPRFQSIKCPISPEGLKAQAAVQVQW 220
 QY 186 VTRKGYPLAK-----LGNTRAA-----VNPDTATVDLNVVVDSCRPIAF 224
 Db 221 YADNGYNLARVLSIEPNRQGLININVAEGLVSDIKFRFVNDGDKTIDSN-----GNPV-- 273
 QY 225 GDFEITGQRYPEQIVSGLARFQPGPYDIDLDDFQQALEQNGHYSGASQAQDFRLQOG 284
 Db 274 -----GCRKPDFLROOL- KLQPGQVQFQENIVKQDVQQLYRTGLFQSVNFAFADATKL 326
 QY 285 DRVPKVSYTEVRHKLETGIRLDSYGLGGKIADYYNLFNK----- 327
 Db 327 DMI---YELKENGARALNIGSYNGDVGMLGTINQQDQIGGKNDTLIANVGLSRTDLQF 383
 QY 328 -----GY-----IGSVYMDMKY-----ETTLAAGISQPNRYR 355
 Db 384 DTKEISPYRNTSDRLCYTVNAFRREISETDEDEIKLANGDKVREGKVGTSISLQPID 443
 QY 356 GNYWTSNVSYNRSSTQNLEKRAFSGGIWVRDAGTDARLGAFLAERKTPGSDIDLGN 415
 Db 444 G--WNASLGNYRST-----IRQKQNIITPTDAQ--GNPLSVSGTGVD--- 483
 QY 416 SHAPMLTASMKROLLNVLHPENGHYLDGKIGTTLTGLTFLSSTALIRTSARAGYF-FTPEN 474
 Db 484 -DLTVTSFATKQDRNPINPTQGSVV--RVSTEQSVPIQCGNISMNRLKADYSQYVPVN 540
 QY 475 ----KKLGTTLIRGOAGYVARDNADVPGLMFRSGGASSVRGYELDSIGLAGPNSGVLP 530
 Db 541 IPNSQTPOVFNLVQAGTVL----GNLPYPTFNILGGSNVRGYDAGVHGG----- 588
 QY 531 ERALLVGSLEYQLPFTRTLSGAVFHDM-----GDAA-ANFKRMKLKHGSGI-----GVWR 579
 Db 589 -RSVLASAERYEPPIVPIVGGVLADFADFASDILGSGDVLGNPAGVRGKPGSGYGCAGIRV 647
 QY 580 FSLAPLSPFDIAVGHSDK-KIRWHISLGTFRF 609
 Db 648 DSPGLGIRAD--YGINDDGESRVHLIGQRF 676
 RESULT 28
 F87486
 outer membrane protein CC1915 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: F87486
 R:Nierman, W.C.; Feidblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Held
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fri
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; UID:21173698; PMID:11259647
 A:Accession: F87486
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-769 <STO>
 A:Cross-references: GB:A0005673; NID:g13423368; PIDN:AAK23890.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC1915

```
Query Match          6.6%; Score 210.5; DB 2; Length 769;  
Best Local Similarity 22.1%; Pred. No. 4.3e-06;  
Matches 145; Conservative 82; Mismatches 247; Indels 183; Gaps 29  
  
QY      61 DSEIKDWVEEHLPLTQQOEEVLDKEQTGFLEAEAPDN-----VKTWLRSKGY 108  
       |::|::||::|::|::|::|::|::|::|::|::|::|::|:  
Db      187 DNDLRDV-----IVTKESRWYKILTSNODYPDRIEYDRQLRKHYNRGY 232  
               |::|::|::|::|::|::|::|::|::|::|::|:  
QY     109 FSSKV--SLTE----KDG-AYTVTHTTPGPRTKIANGVGAILGDILTSDLNLAAYVRALEN 161  
           ||::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 233 FDFRVISSVAELAPDKNGFAVYTTLEEGPKYRFGKITVETELKKL-DGNLLAQI----- 285
Qy 152 WQOPV--GSPFDODSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNPDTATVDLNVVDSG 219
Db 286 --LVRRTGQLYEDERIDYDAITFAAGAAFAFDV-RPRVYPNRETKTVDVYFQVREG 342
Qy 220 RPIAFGFEITGTQRYPEQIVSGLARFQPGTPTDLDLLDFOQALEQNGHYSGASVQADF 279
Db 343 PRVYDRIDVGNTRLDVYLRLEVAESDAYNRVLDRSKNNMRRLGFFKEVEIE-DA 401
Qy 280 DRLOGDRPVKVSVEVKRHKLETGRDSEYGLGKIAIDYNNLFNKGITGYSVVDMDK 339
Db 402 PGSAPDRTSLRVKEOPTGELS-----FSAGYS-----SIDK 434
Qy 340 YETTLAAGISOPRNYRNYWTSNYSNRSTTONLEKRAFSGGIWYVRDRAGIDAR-LG-- 396
Db 435 --LVLDVGITE-RNFRG-----GQNLRARASVGLRQIDFGFSEPRFLGRN 479
Qy 397 -----AFLBGRKIPGSDIDLG-----NSHATMLTA 423
Db 480 LVAGVNLTYFRDYLSEFAAYDTKSVGQDVRFGFPLTNDSSMSLRYTVRQDEVSADSLCA 539
Qy 424 SWKRQLNNVLPENHGYLDKIG-----TTLGTFL-----SSTALI 460
Db 540 SGS---VSQILQORGAYITSLIGYGLRIDKRNDRPINTPRGWFADLNODLAGVGDVKYL 596
Qy 461 RTSARAGYF--FTPENKLGITPIIRQAGYTVARNADVPGLMRSGASSVRGYELDS 518
Db 597 KTEADAQWYMGFT---KDL-VFSATGSFGYIEGWGDNYRINDREYRGG-TSFRGFEIAG 651
Qy 519 IGLAGPNSVLPERALVGSLEQY-----LPFTRTLSCAVPHDMGDA-- 561
Db 652 IGRPDISSFNMSCAKLYAISTELAVPFLPEOYGKAALEFSDVGTAGLLDDVDROSP 711
Qy 562 ----ANFK-RMKLHGSLGVRFSPFLAPSFSDIAYGHS-----DKKIRWHISLGRF 609
Db 712 GVFPNTKDNGLRASAGISIDMKSPGPIREDFISRLSKEDYDRDTETTFSTSTRE 768

RESULT 29
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C.Species: Haemophilus influenzae
A.Variety: type b
C.Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C.Accession: JC4078
R.Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A.Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae
A.Reference number: JC4078; MUID:95255676; PMID:7737523
A.Accession: JC4078
A.Molecule type: DNA
A.Residues: 1-797 <FLA>
A.Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:9537448
A.Experimental source: type b
C.Superfamily: protective surface antigen D-15
C.Keywords: surface antigen
F.1-19/Domain: signal sequence #status predicted <SIG>
F.20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 6.6%; Score 209.5; DB 2; Length 797;
Best Local Similarity 20.0%; Pred. No. 5.3e-06;
Matches 130; Conservative 103; Mismatches 280; Indels 137; Gaps 24;

Qy 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGLAEAPDNVKTMLRSKGYFSSKVSLT----- 116
Db 184 ESVSSSTLQEQWELQPSDNNKGNKFNKFEQAQEKDLSQIRDYVLLNNGYAKAKITKTDVOL 243
Qy 117 --EKDGA-YTVHITPGPRTK-----IANYG--VAIIGDILSDGNLAERYRNALENWOOP 165
Db 244 NDEKTKVNTVDVNEGLQYDLRSARIIGNLGGMSAELEPLLSALHLNDTFRRS----- 296
Qy 166 VGSDFDQDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNP-----TATVDLNVVDSGRP 221

Db 297 -----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 343
Qy 222 IAFGDFEITGTQRYPEQIVSGLARFQPGTPTDLDLLDFOQALEQNGHYSGASVQADFDR 281
Db 344 LTVRQLRFAEGNTVSADSTLRQEMRQOEGTWYNSQLVELGKIRLDRGTGFFE--TVENRTDP 401
Qy 282 LQG--DRVPKVSVEVKRHKLETGRDSEYGLGKIAIDYNNLFNKGITGYSVVDMDK 339
Db 402 INGSNDEVDVYKVKERTGTSINFGYGTGTSIGSIYQASVVKQDNFLCTGAAVSTAGTKND 461
Qy 340 YETTLAAGISOPR-----NYRNYWTSNYSNRSTTONLEKRAFSGGI---WYVRDRA 389
Db 462 YCTSVNLGITYPEYTKDGVSLGGNVFFENYDNSKSDTSNYSRRTTYGNSVTLGFFPVNEN 521
Qy 390 GIDARLG-----AEFLAG-----CRKIPGSDIDLGNHSHATMLTASWKROL 430
Db 522 SYIYVGLGHTYKNTSFALEYNRNLYIQSMKFKGNGIKTNDFD-----PSFGWNYNSL 573
Qy 431 NNVLHPENHGYLDKIGTTLGTLSSTALIRTSARAGYFTFPENKLGITPIIRQAGYTV 490
Db 574 NRGYFPPTKG--VKASLGGRTVTPGSDNKYKLSADVQGFYPLDRHLLWVSAKASAGYAN 631
Qy 491 ARDNADVPGLMRSGASSVRGYELDSIGLAGPN-----GSLVPE 531
Db 632 GFGNKRLPFYQTYTAGSILGRFAYGSI---GPNAIYAEYNGSGTGTFFKISSDVIGG 688
Qy 532 RALLVGSLEYOLP-----FTRLSCAVPHDM-----GDAAANFKRM----- 567
Db 689 NAIATASAEILVPTPFVSDKTSQNTVTRSLFVDAASVYNTKWSKNGLESYVLRKLDPYG 748
Qy 568 ---KLKHGSGLVGWESPLAPSFSDIA-----YGHSDKKIRWHISLGRF 609
Db 749 KSSRIRASTGVGFQWQSPIGPLVFSYAKPIKKYENDDVE-QFQFSLGGSF 797

RESULT 30
F64102
protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20)
C.Species: Haemophilus influenzae
C.Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
C.Accession: F64102
R.Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidma
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A.Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent
A.Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A.Reference number: A64000; MUID:95350630; PMID:7542800
A.Accession: F64102
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-808 <TIGR>
A.Cross-references: GB:L42023; TIGR:HI0917
C.Superfamily: protective surface antigen D-15
C.Keywords: surface antigen

Query Match 6.6%; Score 209.5; DB 2; Length 808;
Best Local Similarity 20.0%; Pred. No. 5.4e-06;
Matches 130; Conservative 103; Mismatches 277; Indels 141; Gaps 24;

Qy 61 DSEIKDMVEEHLPLITQQOEVLDEKQETGLAEAPDNVKTMLRSKGYFSSKVSLT----- 116
Db 197 ESVSSSTLQEQWELQPSDNNKGNKFNKFEQAQEKDLSQIRDYVLLNNGYAKAKITKTDVOL 256
Qy 117 --EKDGA-YTVHITPGPRTK-----IANYG--VAIIGDILSDGNLAERYRNALENWOOP 165
Db 257 NDEKTKVNTVDVNEGLQYDLRSARIIGNLGGMSAELEPLLSALHLNDTFRRS----- 309
Qy 166 VGSDFDQDSWNSKTSVLGAVTRKGYPLAKLGNTRAAVNP-----TATVDLNVVDSGRP 221
Db 310 ----DIADVENAIKAKLG---ERGYGSATVNSV-----PDFDDANKTLAITLVVDAGR 356

Qy 222 I A G C D E I T G T Q R Y P E Q I V T S G L A R P Q G T P Y D I L D L L D Q Q A L E Q N G H Y S G A S V Q A D F D R 281

Db 357 L T V R Q L R F E G N T V S A D S T L R O E M R Q Q G T W Y N S Q L V E L G K I R L D T G T F F E - T V E N R I D P 414

Qy 282 L O G - - D R V P K V S Y T E V K R H K L E T G I R L D S E Y G L G K I A Y D Y N L F N K G Y I G S V V M D M K 339

Db 415 I N G S D E V D V Y K V K E R N T G S I N F G I C Y G T E S G I S Q A S V K O D N F L G T G A A V S T A G T K N D 474

Qy 340 Y E T T L A A G T S Q R P - - - - - N Y R G M Y T S V N S Y N R S - T T Q N L E K R A F S G G I - - - W Y V R D R A 389

Db 475 Y G T S V N L G Y T E P Y F T K G V S L G G N V F P Y D N S K S D T S N Y K R T T Y G S N V T L G F P P V N E N N 534

Qy 390 G I D A R L G - - - - - A E F L A E - - - - - C R K T P G S D I D L G N S H A T M L T A S K R Q L L 430

Db 535 S Y Y V G L G H T Y N I S N F A L E Y N R N I Y T Q S M K F G N G I K Y N D F D - - - - - F S F G W N Y N S L 586

Qy 431 N N Y L H P E N G H Y L D G K I G T T L G T F L S T A L I R T S A R A G Y F F T P E N K K L G T F T I R Q A G Y T V 490

Db 587 N R C Y F P T K G - - V K A S I G S G R V T I P G S D N K Y K L S A D V Q G F Y P L D R H L W V S A K A S A G A N 644

Qy 491 A R N A D V P S C L M F R S G G A S V R G Y E L D S T G L A C P N - - - - - G S V L P E R A 533

Db 645 G F G N K R L P Y Q T Y T A G G I S G L R G F A Y G S I - - - G P A I Y A E H G N G T F K K I S S D V I G G N A 701

Qy 534 L I L V G S L E Y O L P - - - - - F T R T L S G A V F H D M G D A A N F - - - - - 564

Db 702 I T T A S A E L I V P P F V S D K S Q N T Y T S L F - - - V D A S V N T K W K S D K S G L D N N V L K S L P D Y 758

Qy 565 - K R M K L U H G S G L V G W P S P L A P S F D T A - - - - - Y G H S D K K I R N H I S L G T R F 609

Db 759 G K S S R I R A S T G V G F Q W O S I P L V F S Y A K P I K Y E N D D V E - Q F Q F S I G G S F 808

RESULT 31

A81430

outer membrane protein Cj0129c [Imported] - Campylobacter jejuni (strain NC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2000

C:Accession: A81430

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Bashan

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitene

Nature 403, 665-669, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni

A:Reference number: A81250; PMID:20150912; PMID:10688204

A:Accession: A81430

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-739 <PAR>

A:Cross-references: GB:ALL39074; GB:AL111168; NID:96967505; PIDN:CAB72613.1

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0129c

```

Query Match      6.5%; Score 206; DB 2; Length 739;
Best Local Similarity 16.3%; Pred. No. 8.1e-06;
Matches 140; Conservative 102; Mismatches 312; Indels 210; Gaps
QY 19 AYAPAAADLSEKKAAGAFALFNKSP-DTESVKLKPKP-PRVIDT-----59
DB 13 AMANAATKDKIFGLNHLNNTSAINIAGLKKGEEINPAKINTAILNLYKQNTFENIAVE 72
QY 60 QDSEIKDWVEEHLPLIT-----QQQEEVLDRKEQTGF LAE----EAPDNVKTML 103
DB 73 NNGGLEIIVTEKETPIAKVTITGIASNDRKQVESILGKRGCTLLDEGNIKEATERIKAY 132
QY 104 RKGYFSSKVSLETKDQAYTVHTTPPRTK-IANYGVAILGDILSDGNLAEYVRNALENN 162
DB 133 EAKSYFDITVEYKKK-----TLENTDGLGLEFIVNNGENIIIDNVHLSGAKKFSYSDIEPA 188
QY 163 QOPVGSDFDQDSW-----ENSKTSVLGATVRKGYPLAKLGNTRAAVNPDTATV 210
DB 189 VYNKSEKPEPMWNGNRDCKLVKPFELNSDSRSRLADEYMKKGYLDVVOVSSPKLYITDTYOA 248

```

[illegible]

Query Match 6.48; Score 202.5; DB 2; Length 861;

Best Local Similarity 20.4%; Pred. No. 1.8e-05;
Matches 131; Conservative 77; Mismatches 254;
Indels 179; Gaps 23;

[illegible]

QY 87 O-TGFLAEAPDNVKTMLRSLKGYF-----SSKVSLT-EKGAY-TVHIITPGPTKIANGV 138
:
Db 224 KLSGDLE----EKLNWSYLDRGYDFNDTSQVSISPEKNMFITAGVTGGDYKISSIK 278
:
QY 139 VAILGDILSDGNLAERYALNENMOOP-VGSDFDQDDSWENSKTSVLGAIVTRKYGLAKLIG 197
:
Db 279 VT-----GNTV-LPQBKEIKLVPIKGTDFISRVLLYESAAIINTLSNICGYAFSKV- 328
:
QY 198 NTRAAYNPOTATVDLNVVVDSGRPTAFQDGFEITQTORYPQIVSGLGARFOPGTPDLDLL 257
:
Db 329 NPITPNRADRTVAVNLHVIPGPVRTVYHQILFKGNTRTSDDEVLRREKROFENSWSOQA 388
:
QY 258 LDFQQALEONGHYSGASVQADFDRLQGDRVPVKVSVTEVRKHRLKETGIRLDSYGLGGKI 317
:
Db 389 DRSKIQLQRLGYFEADVDESTPVPCGSNDQDVIVTVTKETTSFGSFQVGLGYSKTYGVTTSV 448
:
QY 318 AYDYYNLFNKKYGTSVWDMDKYETTLAGISQPRNRGNWTSN-----VSY 365
:
Db 449 QLSQNPLGSGNRVSVDAISRYSQRYFSFTNP-----FFTDNGVSLGYNLAYOKLDY 502
:
QY 366 NRSTTONLEKRAFSG---GI-----WVVRDRACIDARLCAEFLAEGRKIPGSD-- 410
:
Db 503 SDENAAYNSKRMSGQTGFIPITENDTVSMVI-----GASNQITTF-----PGSTPK 551
:
QY 411 --TDLGNS-----HATMLTASWKROLNNLVHPENGHY-----LDG-----KIGT 448
:
Db 552 AIIDYIDAVGOQRTFRAMTELGCWARDTRNFNLMGYORIGAENVTLPGSTTIKYIKIN 611
:
QY 449 TLGTF--LSTALIRTSARAGYFTFPENKKILGTFIIIRGOGYTVARDNADVPSSL-----M 502
:
Db 612 QISKYWPPIPALVNLVLTREVGY-----GDDYCKSNTRLIPDGTVA-----TASGLPF 661
:
QY 503 FRSGGASSVRGVELDISGLAGP-----NGSVLPERALLVGSLSE-----YOLPPFRTLS 550
:
Db 562 FYAGTNSVVRGFRDNTLGPRESVTALYVQGOPLGSGSPKTVGSTEMFWPKLFUDSPSARISA 721
:
QY 551 GAVFHDWGDAANFRKMKLKHGSGLGVRWFSP LAP-----FSFDIAYHSKDKIRWHISLG 606
:
Db 722 FLDFGVNFVGNFNNEFRKANELRASSGVALLWRAPIGPISYAFAPIKKNENDEIERLQTF 781
:
QY 607 TRF 609 -
:
Db 782 GF 784

RESULT 35
AH3355
outer membrane protein BMEI0830 [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AH3355
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mufier, C.; Los, T.;
Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.
Proc. Natl. Acad. Sci. U.S.A. 95, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella
melitensis strain 16M
A:Reference number: AD3252; PMID:11756688
A:Accession: AH3355
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1781 <KUP>
A:Cross-references: GB:AE008917; PID:N:AAL52011.1; PID:g17982775; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0830
A:Map position: I

```
Query Match      6.1%   Score 194.5; DB 2; Length 781;
Best Local Similarity 19.5%; pred.No. 5.2e-05;
Matches 141; Conservative 115; Mismatches 284; Indels 183; Gaps 30;
```

Db 103 VKERSVVNVFLQGNKKTKDDPLARAVOLKPRAPFDMATMEAD-----K 140

Qy 81 EVLDKBOTGFLAEAAPDNVKTKMLRSKCYSSKYSLTEKDGYATVHITPGPRTKIANVGVA 140

Db 147 EAIKAAAYSHIGRSADTAVNARTVDLQGC-----RVNV-----VVEINSGSTKIANIEFV 195

Qy 141 -----ILGDILS-----DGNLAEYRNALENWQOPVGSDFDDOSWENSKTSVLGAVTRK 189

Db 196 GNOAFSGRRRLRDVISTKRSNPLSWLTRDV-----YDEGLQADEETLRFFYNNR 245

Qy 190 GYPLAKLGNTRAANVPDITAVDLNVVYDSGRPIAFGDFEITGT-QRYPEQIVSGLARFPQ 248

Db 246 GYADFVRLSSNAVLDPSTNEYTITITVDEGPRTYFEGDVSVESVDGVDQTQALDRLVKTRT 305

Qy 249 GTPYDL-----DLLLFOQALQNGHYSGASVOADEDR-LQGDVRPVKVSUYEVKHKLE- 302

Db 306 GKPYSAKETDSVLVSTVESAGS-YAFAKVPEPRGDNRNFENITISVYVSDQGPRAYIOR 364

Qy 303 ---TGIRLDSYGLGCKIAYDYVNLFNKGYIGSV---VMDMKYET---TLAAGISOPR- 352

Db 365 IEIRGNDKTRDYVIRREFDLNEGDAFNOVHVQAKRRLEVLDFQTVNISTAPG-SEPDQ 423

Qy 353 -----NYRCNWTSNVSYNRSTTQNLKRAFSGGIWYVDRPAG----- 390

Db 424 VILVDVVEKSTGERSIGGGYTTGCGSPGAQVEAAITERNFLRGQYIRISAGAGQDDMR 483

Qy 391 -----IDARLGAELAEGRK-----IPGSDI----- 411

Db 484 NYGLSTPEVFLGYRLSAGFOVPRRSYRVNDDYDQVTGTTIRFGLPTIDNFSAGIAYSIL 543

Qy 412 -----DLNGSHATML-----TASKRR-----QLLNVLHPENGHYLQCKIG 447

Db 544 VQEKYOLFPGDAENYYPALAEANSPWLRSVSVSYLTSYSSIDDKNPHDGLY--GKFI 601

Qy 448 TTLGTFLSALTIRTSARAGYFFTPENKKLGTFTIRQAGYTVARDNADVPGLMFRSGG 507

Db 602 QEFAGLGGDAAKVKVTKTEKNGYIYQTLSEADIVGLLGVAGGYTHEFGDDGVRFIFOLPKN-S 660

Qy 508 ASSVRGYELDSIG--LAGPNCG--VLPERALLLVGSLEYO----LPFTRTLSGAVF---- 554

Db 661 SDIIRGFKFNGIGPYQDAKNCKRYWMGGTTYFSGTAEVQFPMVLPESLGVRGAFADAA 720

Qy 555 -----HDMGDAAANFKMKLHGSLGVYRWFSPLAPSFSDIAY---GHSDDKKIRWHISLG 606

Db 721 TLXGNDTPDISGDDR--KLRSAGVSIAMWASPPGLRFDYAPPAKADTDKVOFNFGVS 778

Qy 607 TRF 609

Db 779 TKF 781

RESULT 36

D97527

cmpl protein precursor (U51683) [imported] - Agrobacterium tumefaciens (str. D97527)

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2001

C:Accession: D97527

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C. Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens (Strain D97527)

A:Reference number: A97359; PMID: 11743194

A:Accession: D97527
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PTD:gI5156447; GSPDB:GN0000000001
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match		5.8%	Score 185;	DB 2:	Length 774;
Best Local Similarity		20.9%;	Pred.	No.	0.00022;

Db 167 VSKDANISGLN-----DQATTVOVS-----TKDQFNSRFTSPYRLG 205
Qy 371 -----QNLKRAFGSGGIWTVRRDAGIDARLGAEPFL 400
Db 206 OPERLGYSVTAFRDRQISRTEDDEIKLANGNTRREGFGGSAVLRSGFDENAAALGLNYT 265
Qy 401 -----AEGRKIPGSDIDLGNSHATMLTASMKROLLNNVLHPENGHYL-----D 443
Db 266 RISLRDKDYNVRADGLGNPLSVSGKGIDDLFTLSLATVRLDRHLNPTQGSLLILSTE 325
Qy 444 GKIGTTTGLTSLST-----ALIRTSARAGYFTTPE-----NKKLGTFTIIRGOAGYTVA 491
Db 326 QAIPLGLGNIIFSRRFGNYIQVLPVTWIGNQNQLTDNPMVAVNLQGTIF----- 375
Qy 492 RDNAADVPSGLMFGSGASSVRGVELDSSIGLAGPNSVLPERALLVGSLEYQLPRTPLSG 551
Db 376 ---GEPPADAFNLGGLNSRGV-----GSKLASGRSYGLASVEFRFPVIESIGG 423
Qy 552 AVFIDMGDAANFKR-----MKLKHGS-----GLGVRWFSPAPFPFDIAYGHSK-KI 599
Db 424 VVFTDFASDFGSGKTVIGEGVVRNKPCKSGFGYGLGVRLNSPFGFLFRGDL--GVSDQGEV 481
Qy 600 RWHISLQTRF 609
Db 482 RFEITTCORF 491

RESULT 39
F71910
probable outer membrane protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: F71910
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-906<ARN>
A:Cross-references: GB:AE001492; GB:AE001439; NID:94155147; PID:9415515
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0600

Query Match 5.5%; Score 174; DB 2; Length 906;
Best Local Similarity 21.0%; Pred. No. 0.0016;
Matches 123; Conservative 72; Mismatches 202; Indels 188; Gaps 23;

Qy 186 VTRKGYPLAKLGNTRAVNPDTATVDLNVVV-----DSGRPIAFGFEITGTQRYPEQI 239
Db 349 IADKGAF-----AVVAPDLDDKDEKGLVKVIYRIEVDGMVHINDVIISGNQRTSDRI 401
Qy 240 VSGLARFPQCPGYDLDLLDFOALEQNGHYSGASVQADFDRLQGDRAVPVKVSVTEVKRH 299
Db 402 IRRELLGPKDKYNLTKLNSNSLRUGFFS--KVTEEKRVNSLLMDLLYSVEEGRTG 459
Qy 300 KLETGRILDSEYGLGKIAIDYNNFNKGYIGSVVWD-----MOKYETTLAAG- 347
Db 460 OLOPGLGYSGYGLMLNGSVSERNLFGTQOSMSLYANTATGGRSYPCMPKPGAGRPAGN 519
Qy 348 --ISQPR-----NVR-----GNWTNSVNSKSTQ 371
Db 520 LSLTNPRIFDSWYSSTINLYADYRISYQIQGGGFGVNVGRMLGNRTHVSLCYNLNVTK 579
Qy 372 NLEKRAFGSGGIWTVRRDAGID-----ARLGAEFFLAERK--IPGSDIDL----- 414
Db 580 LL---GFSSPL--YNRYSSVNEFASPROCSPTASVNIINRLSGGRTPLPVPSCSPGAITT 635

Qy 415 -----NSHATMLTASMKRQL-----LNNVLHPENG-----HYLDGKIGITLGTFLSST 457
Db 636 SPEIKGIWDRDYHTPTTSSTFLDVSYDNTDDYFPRNVISSYATWSGLPSSGTLNSWN 695
Qy 458 AL-----IRTSARAGYFTTPEKNKLGTFI-----IRGAGYTVARNAD--VPSGLMFRSGG 507
Db 696 GLGNNVRNTKVYCKFAAYHHLQKLLIDLIARFKTQGY-IFRYNTDDYLPANSTPYMGG 754
Qy 508 ASSVRGYELDSIGLAGPNSGVLP-----CHSDKKIRWHISLQTRF 509
Db 755 VTVVRGPR-----NGSITPKDEEGLMGGDGIPTASTELSYGVLKAAMRLAWFFD 805
Qy 532 -----RALLVGSLEYQLPRTPLTSCAVPHDMGDAAANFKRMKLGHSGLGVRWFSPPLAP 585
Db 806 FGELTFKTPTRGSEFFYNAPTTT-----ANFKDYGVVAGFERATWRASTGLQIEWISHPMG 861
Qy 586 --PSFDAY-----CHSDKKIRWHISLQTRF 509
Db 862 LVLIFFIAFFNQMGDNGKCKGLCFNPMNDYTOHFEFSMGTRF 906

RESULT 40
D86528
omp85 analog [imported] - Chlamydothila pneumoniae (strain J138)
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: D86528
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: D86528
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-790<STO>
A:Cross-references: GB:BA000008; NID:88978674; PID:BA098510.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: yaet

Query Match 5.5%; Score 173.5; DB 2; Length 790;
Best Local Similarity 20.7%; Pred. No. 0.0014;
Matches 148; Conservative 92; Mismatches 266; Indels 209; Gaps 33;

Qy 22 PAADLSNKAAFGALPKNKPSTESVKLPKPPVRIQDQDSEIKDMVEEHLPLITQO--Q 79
Db 87 PKVEFSEGG--TNIALHLIAKPSIRNIHISG-----NQVVPEHKILKTLQIYR 132
Qy 80 EEVLDKTEOTGFLAEAPDNVKTMLRSKGYFSSKVSLETEKCAVTVH-----ITPGPRTK 133
Db 133 NDIFERK--FL--KGLDRLRYLLKRGYFASVDYSLEHNOEKGHIDVLKINEGPCGK 188
Qy 134 IANVGVAIG-----DILSDGNLA-----EY 155
Db 189 IKOLTFSGISRSEKSDIQEFTQIKQHTSTTSFTGAGLYHPDIVEQDSLAITNYLHNNGY 248
Qy 156 RNALENMQQPVGSDFDQDSWEN-----SKTSVLCAVTRKGYPL----- 193
Db 249 ADAI-----VNSHYDLDDKGNILLYMDIDRGSRYTLGHVHIQGFVLPKRLEKOSQVG 302
Qy 194 -----AKLG--NTRAAV-----NPDTATVDLNVVDSGRPIAFGD 236
Db 303 PNOLYCPDKTWDGAHKIKQYAKYGYINTNVNDVLPFHATRPYDVTVYEVSESPYKVGL 362
Qy 227 FEITGTQRYPEQIVSGLARFQPTPYDLDLLDFOALEQNGHYSGASV-----QA 277
Db 363 IKITGNHTKSDVILHETSLFPDGTFRNLKLEDTEORLNRNTGYFQSVSVTVRSQOLDPMG 422
Qy 278 DFDRLQGDRAVPVKVSVTE-----VVRHKEGTGIRL--DSEYGLGKIAIDYNNFNK 338
Db 423 NADQYRQIFVEVRETTTGNLGLFGFSSLDNLFGGIELSESNFDLFGA-----RNIFSKG 477
Qy 329 Y-----IGSVVMDMKYETTLAAGISOPRNYRGNWTWSNV-----SYNRST 369

$$\frac{21^2}{2}$$

$$\frac{21.8}{43.6}$$